应用随机过程教程

– 及其在算法和智能计算中的应用

龚光鲁　钱敏平 著

清华大学出版社

龚光鲁，钱敏平著 **应用随机过程教程及其在算法与智能计算中的应用**

清华大学出版社,2003

内容提要

本书概述了应用随机过程的基本内容以及近代的重要进展与重要方法.我们并不假定读者具有测度论的知识.在更多地使用不严格的推理时，我们尽最大的努力做到理论与算法兼顾.

第 1 **章是概率论与数理统计的复习.第** 2 **章典型分布的随机数的生成方法**, **第** 3 **章** Poisson **过程**, Brown **运动与随机徘徊.第** 4 **章介绍更新过程.第** 5 **章是离散时间的** Markov **链**, **得到的** Markov **链的遍历性定理**, **是与** Markov **链的初始状态无关的**, **这样就满足了统计物理中的"各态历经性"的要求.第** 6 **章连续时间的** Markov**.第** 7 **章介绍排队理论梗概.第** 8 **章致力于在实际应用中有力的随机模拟方法**, **即** Markov Monte Carlo **方法的基本原理.第** 9 **章介绍以图像处理为背景的随机场与用随机迭代系统方法处理图像的方法.第** 10 **章是隐** Markov **模型**, **这是近年来强有力的建模工具.第** 11 **章的内容为** Gauss **系，更多的内容是时间序列的常用模型.第** 12 **章** Markov **过程**, **鞅论**, **随机微分方程与扩散过程，包括随机微分方程的数值近似解法.第** 13 **章介绍金融数学中的证券模型与其衍生金融工具的定价.第** 14 **章扼要地介绍保险中的集体风险理论.第** 15 **章中的算法**, **包括** EM **算法**, **人工神经网络**, **遗传算法与自组织算法.第** 16 **章为离散时间的** Markov **决策的梗概.第** 17 **章介绍** Poisson **随机微积分与自激点过程.**

本书适合于作大学工科, **理科非数学专业**, **应用数学的非概率统计方向**, **医学**, **心理**, **经济金融等本科高年级学生和研究生的教材或参考书；也是教师**, **研究人员**, **工程师**, **设计师以及使用应用随机过程分析数据资料工作者的重要参考书.**

龚光鲁,钱敏平著　应用随机过程教程 – 与在算法和智能计算中的应用

清华大学出版社,2004

<应用随机过程教程– 及其在算法和智能计算中的应用> 前言

作者在 5 年前向读者奉献了一本教科书<应用随机过程>，它是由北京大学出版社出版的．那本书主要是为理科, 特别是为数学与概率统计专业学生而设计的. 该书与当时市面上的应用随机过程教科书相比, 其长处在于着重强调随机过程在各科学领域的新的应用, 提供了不少在应用领域中用随机过程建模的例子．

然而, 当作者之一向工科研究生与高年级本科生授课时, 发现该教材开始所涉及的内容，以及后面章节中的个别内容颇为艰难. 顾及工科, 经济类, 管理类等领域不同读者对象的需要, 我们想到应该更多地扩展与应用方面有关的内容哦内容，模型与概念, 甚至应该介绍一些常用的算法, 即使有些算法不全直接出自随机过程．凡此考虑, 就成为我们再写一本主要为工科设计的 <应用随机过程教程>教材的动机.

本书是主要为广大的工科, 理科非数学专业, 应用数学的非概率统计方向, 医学, 心理, 经济金融等诸多领域的本科生, 研究生, 教师, 研究人员, 工程师, 设计师等撰写的．在本书中，我们并不假定读者知道测度论的知识．鉴于选用本书的对象对概率统计的了解程度与时限很不一致, 根据读者的建议, 我们在本书中将与应用随机过程有关的概率统计的基本要点(但并不完全概括一般概率统计课程所涉的内容), 略去其证明与概念的解释, 写成一个纲要, 作为第 1 章, 以便使用者能作 “在线的” 温习与查阅. 但是, 我们在第 1 章中所引述的内容, 在个别地方会比一般工科概率论的基本课程的要求略为超出一些, 这是为一些有余力的读者准备的, 以便这一部分读者能得到更深一些的领会．我们建议使用本书的教师, 不要把第 1 章放在应用随机过程课程的教学内容中, 而只把它作为学习其他章节时的参考温习材料．

由于我们希望有较宽的适应性, 本书与某些针对个别专门方向应用的应用随机过程教科书相比, 增加了不少内容． 当然我们也舍去了, 或改写了某些经典的内容, 例如关于平稳过程的遍历论, 二阶矩过程与宽平稳过程的谱理论等等，我们基本上并没有展开，而只作了必要的论述．我们冀望使用本书授课的教师和自学的读者, 可以根据学生或读者自身的情况与需要, 把本书的素材分成几个不同的层次, 选用其中的一部分或几个部分, 即尽量不要把本书当作 ”套餐”(menu du jour), 而努力把它作为 ”点菜” (a la carte) 的菜单．在使用本书学习时, 要学会熟悉本书的目录与符号表，经常浏览并翻阅本书最后所附的名词索引, 以便能更好地使用 “点菜技术”． 需要强调的是, 由于作者希望本书包容一些常见的应用层面, 以及经过简化后的典型应用例子, 并由此安排分析它们的各种工具或算法, 这就使本书的某些内容不可能完全按照一维的次序展开, 也就不可避免地会出现材料使用上的交叉.特别地, 在后面的几章中就明显地出现了这种交叉．这就使 “点菜技术” 更有必要．

本书中带 * 的部分和小字的部分, 只是用作参考与注释, 完全可以略去. 不要因为它们而影响讲授或学习的主线．

本书的撰写原则是，着重与强调想法, 背景与思路．为了使读者更便于理解, 对于命题与定理, 我们只给出简单的证明, 或者直观的证明, 而不追究其严格性．对于非常重要的内

容，例如 Markov 链的与初始值无关的遍历定理，我们不吝冗余，在各种不同的情形反复叙述它的各种形式，以使原来并不熟悉的读者能逐步领会其实质．随着内容的逐步发展，我们也加进了一些对数学背景要求较多的一些概念与论述．对此感到不习惯或者并不需要的读者，完全可以跳过它们，或者代之以更直观的理解．

本书的第 1 章是复习材料．第 2 章是通向实际模拟计算的桥梁．第 3 章，第 5 章，第 6 章和第 12 章是基本理论．第 4 章，第 7 章，第 8 章，第 9 章，第 11 章，第 13 章和第 16 章是应用基本理论．第 10 章和第 15 章是应用中的最常见算法，第 14 章是应用．第１７章则是拓广性的介绍． 更具体地，我们把本书的材料组织如下： 第 1 章是概率论与数理统计复习．第 2 章讲述典型分布的随机数的生成方法，这是 Monte Carlo 算法(即随机模拟算法)的基础．第 3 章阐述随机过程的一般概念，及独立增量过程的重要的例子，包括 Poisson 过程，Brown 运动与离散的随机徘徊． 第 4 章介绍更新过程，它是一种在应用中常见的计数随机过程．在本章中，我们以解释概念为主，较少论及证明．第 5 章是离散时间的 Markov 链．我们从更新序列返回的周期现象来理解 Markov 链的常返态的周期性．在本章中论述 Markov 链的极限性质时，采取了与传统著作中不同的方法，用转移矩阵的平均极限作为支点，避免了细致繁琐的讨论，这一章的侧重点是可逆性和不变分布，用以求非周期的正常返的不可约 Markov 链的平稳分布．由此得到的 Markov 链的遍历性定理是与 Markov 链的初始状态无关的，这就满足了统计物理中的 “各态历经性” 的要求．第 6 章是连续时间的 Markov 链，它在我国学术界有一个特殊的称谓，即所谓Q过程．在实际应用中，连续时间的 Markov 链比离散时间的 Markov 链更为常见．在应用问题中，其转移概率速率矩阵Q是常常可以实际测量到的．在相当一般的条件下(这些条件在实际应用中总是能满足的)，可逆分布和平稳分布所满足的方程都可以由转移速率阵Q表达．这一章的侧重点也是可逆性和不变分布，用以求时间连续的 Markov 链的平稳分布，并得到与 Markov 链的初始状态无关的遍历性定理． 第 7 章介绍排队理论梗概．主要点是了解这类问题的提法与关心的问题．第 8 章阐述 Markov Monte Carlo 方法，其主要功能是对非常高维数的随机向量作取样．着重介绍了 Gibbs 取样法与 Metropolis 取样法及其理论依据．最后讲述了优化的模拟退火的基本思想．第 9 章致力于以图像处理为背景的随机场．在叙述时间离散状态连续的 Markov 链的基础上，简要地介绍了用随机迭代系统方法处理图像．第 10 章论述近代在语音识别，手写体文字识别，DNA 序列信息采挖等实际问题中广有成效的隐 Markov 模型，分析了这种模型的优点与它在应用中的潜力．在叙述用它建模的有效算法时，着重指出其与 EM 算法的联系．第 11 章在论述 Gauss 系的基础上，从应用的角度阐述了时间序列的各种常用的模型，着重于介绍算法．第 12 章是向基本理论的回归，其内容包括 Markov 过程，鞅论浅介，随机微分方程与扩散过程的要义，这是近年来在应用中极为活跃的建模工具．最后还提供了随机微分方程的数值近似解法．第 13 章论及金融数学中的证券模型与其衍生金融工具的定价．特别介绍了二叉模型与随机利率的期限结构．第 14 章着重介绍保险中的集体风险理论． 第 15 章介绍各种算法，主要包括 EM 算法，人工神经网络，遗传算法，Kohonen 自组织算法，适应最小二乘法． 第 16 章给离散时间的 Markov 决策以一个概括性的陈述．第１７章介绍 Ito 随机微积分的推广，主要是 Poisson 过程的随机微积分，它对处理电子学，金融经济学中的一些理论问题，是一个很有用的工具．最后还扼要地介绍自激点过程．

为了查阅方便，本书对定理，定义，引理，命题，例子等，都采用统一的计数系统．

由于本书篇幅较大，我们建议一些初学的基本内容：例如，第３章，第４章的第１节，第３．２段，第４．１段；第５章；第６章；第７章的第１节和第２节，第３．２段；第９章的第２节； 第１０章的第１节和第２节；第１１章的第０节到第４节；第１２章的第１节至第３节，第４．１段．

我们把应用方面的学习内容，留给授课的教师与自学的读者，让他们根据需要选取．

本书主要章节后面的习题是这样配置的： 第 1 章后设置一些复习题．因为第 3 章至第 6 章，第 11 章和第 12 章是主体，所以安排了较多的习题，第 2 章，第 7 章，第 9 章也附有一些习题，而对第８章，第 10 章与第 13 章，则仅有少量习题．此外，第１７章也有一些习题．

本书与我们献给教育界的前一本书一样，同样以强调应用作为宗旨．希望应用随机过程的思想方法，能成为各界朋友在实践与应用中帮助思维有力的工具．我们希望各界朋友能提供反馈的意见与不同的观点，以及适合在教学中运用的材料，以便在本书再版时，能够在各界朋友的帮助下，更上一个台阶．

以应用为背景的随机过程，已经经过了近百年的发展，从各个领域应用中萃取出的典型模型，典型方法，在一本入门的教科书中是难以概括的．况且，哪一些材料，哪一些思想更值得选取，也没有判断标准．作者希望自己的判断尽量与实际接近．对于应用例子，我们也不可能写得很具体，而只指出一些思路，更多的填补与链接，需要读者自己去完成．本书撰写的理念，是以国际前沿为目标的，在这方面，我们更需要各领域的使用者的合作．

清华大学数学科学系 龚光鲁 glgong@math.tsinghua.edu.cn
北京大学数学科学院 钱敏平 qianmp@math.pku.edu.cn

龚光鲁，钱敏平著　应用随机过程教程 – 及其在算法和智能计算中的应用

清华大学出版社,2004

目录

符号说明

1 分量全是 1 的列向量.

$a \vee b, a \wedge b$ a,b 两个数中的大者和小者

a^+, a^- $a^+ = a \vee 0$，$a^- = a \wedge 0$

B_t Brown 运动

$Cov(x,h)$ 两个随机变量 x,h 的协方差

$E_i h = E(h \mid x_0 = i)$, $P_i(A) = P(A \mid x_0 = i)$

$\exp_l$ 参数为 l （均值为 $\frac{1}{l}$ ）的指数分布

I 单位矩阵

$\Gamma(a, l)$ Γ 分布

i.i.d. 独立同分布

$N(m, s^2)$ 正态分布

$U[a,b]$ $[a,b]$ 上的均匀分布

$Ex, Var(x)$ 随机变量 x 的数学期望和方差

$\overset{\Delta}{=}$ 定义为

$x \overset{d}{=} h$ x 与 h 同分布

$x_n \xrightarrow{p} x$ 依概率收敛

$x_n \xrightarrow{d} x$ 依分布收敛

$x_n \xrightarrow{a.e} x$ 随机事件 $\{w : x_n(w) \to x(w)\}$ 的概率为 1

$x_n \xrightarrow{L^2} x$ 均方收敛: $E|x_n - x|^2 \to 0$

d_{ij} Kronecker 记号，$d_{ij} = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j) \end{cases}$

F s 代数(事件体)

I_A 事件 A 的示性函数，$I_A(w) = \begin{cases} 1 & (w \in A) \\ 0 & (w \notin A) \end{cases}$

$\| x \| \overset{\Delta}{=} (x,x)^{\frac{1}{2}} = \sqrt{Ex^2}$

$\Phi(x)$：**在随机变量族**$x \overset{\Delta}{=} \{x_a : a \in I\}$**中任意有限个元素的任意有界连续函数全体组成的集合.**

$\overline{\Phi}(x)$：**包含**$\Phi(x)$**且对** L^2? **收敛性封闭的最小集合.**

L^2：**平方可积的随机变量的空间，**$L^2 = \{x : Ex^2 < \infty\}$

$L(x), \overline{L}(x)$ **为包含** $L(x)$，**且对于均方极限封闭的最小集合**

(Ω, F) **样本空间**

$P(A|B), P(A|h), E(x|h)$ **条件概率，条件期望**

$P(x,A)$ **转移概率**，$P(x,A) = E(x_1 \in I_A | x_0 = x)$

$\mathrm{Proj}_{\overline{\Phi}(x)} h, \mathrm{Proj}_{\overline{L}(x)} h$：$h$ **在空间**$\overline{\Phi}(x), \overline{L}(x)$ **上的投影**

$p_x(x|h=y)$**或** $p_{x|h}(x|y)$ **条件密度**

$\mathrm{Poisson}_l$ **参数为** l **的** Poisson **分布**

r_{xh} **随机变量**x, h **的(线性)相关系数**

$R(k)$ **平稳随机系列的相关序列，** $R(k) = E(x_n x_{n+k})$

R^d, R d **维 Euclid 空间，欧氏直线**

s^2 **样本标准差**

$\vec{x}, (\vec{x})^T$**，**$\vec{x} \le \vec{y}$ **列向量，其转置，所有的分量都不大于**

Z^d, Z d **维格点**，**一维格点**

参考文献

[A] Arnold, L. Stochastic Differential Equations, Theory and Application. John Wiley & Sons, 1974

[AC] **安鸿志**, **陈敏**, **非线性时间序列分析**, **上海科技出版社**, 1998.

[As] Asmussen, S. Applied Probability and Queues. Wiley, 1987.

[B] Bremaud, P. Markov Chains, Gibbs Fields Monte Carlo Simulation, and Queues. Text in Applied Mathematics, 31. Springer, 1999.

[BT] Box, G. E. P. and Tiao, G. C. , Bayes Inference in Statistical Analysis. Wiley-Interscience, 1992.

［BMP] Benveniste, A. Metivier, M. Priouret, P. , Adaptive algorithm and stochastic approximations, Berlin: Springer, 1990.

[BW] Bhattacharya, R. M. and Waimire, W. C. , Stochastic Processes with Applications. John Wiley & Sons, 1990.

[C] **钟开莱**, **初等概率论附随机过程**. **人民教育出版社**, 1980.

[CB] Casella, G. Berger, R. L., Statistical Inference. 2nd Edition. Duxbury Advanced Series. 2001.

[CT] Caudhry, M. L. Templeton, J. G. C., A First Course in Bulk Queues. John Wiley & Sons, 1983.

[CoT] Cover, T. M. and homas, J. A. Elements of Information Theory. John Wiley and Sons, 1991.

[D] Dempster , A. P. Laird, N. M. and Rubin, D. B. , Maximum Likelihood for incomplete data via the EM Algorithm. Journal of the Royal Statistical Society B, 39:1-38, 1977.,

[De] Devroye, L., Non-Uniform Random Variate Generation. Springer-Verlag. New York, 1986.

[E] Edwards, A. W. F., Likelihood. Johns Hopkins University Press, 1992.

[ET] Efron, B. and Tibshirani, R. J., An introduction to the Bootstrap. Chapman and Hall, 1993.

[G] **汉斯** U. **盖伯**, **数学风险论导引** (**成世学**, **严颖 译**), **世界图书出版公司**, 1997.

[H] **何声武**, **随机过程论**. **华东师范大学出版社**, 1989.

[H-LL] Hernandez-Lerma, O. and Lasserre, J. B., Discrete-Time Markov Control Processes, Basic Optimality Criteria. Springer, 1995.

[IS] Isaacson, D. L. and Madsen, R. W.,Markov Chains Theory and Applications. Wiley-Interscience , John Wiley & Sons, 1976.

[L] **陆懋祖**, **高等时间序列经济计量学**, **上海人民出版社**, 1999.

[Li] van Lieshout, M., Markov Point Processes and their Applications, Imperial College Press, The Nietherland, 2000. (distributed by World Scientific Publishing Co, Singapore).

[LK] Law, A.M. and Kelton, W. D., Simulation modeling and Analysis. Mcgraw-Hill, 1991.

[MR] Meng, X. L. and Rubin, D. B., Recent Extension to the EM Algorithm. Bayesian Statistics, 4, 307-320,1992.

[N] Neal, R. M., Bayesian Learning in Neural Networks. Springer, Lecture Notes in Statistics, 1996.

[Ne] Salih N. Neftci, An Introduction to the Mathematics of Financial Derivatives. Academic Press, 2nd edition, 2000

[NH] Neal, R. M. and Hinton, G. E., A new View of the EM Algorithm that justifies incremental and other Variants. ftp://archive.cis.chio-state.edu/pub/neuroprose/neal.em.ps.Z

[QG] **钱敏平**, **龚光鲁**, **应用随机过程**. **北京大学出版社**, 1998.

[R] Rabiner, L. R., A Tutorial on Hidden Markov Models and Selected Application in Speech Recognition, Proc.IEEE,77(2),257-286, 1989.

[Ro] Ross, S., Stochastic Processes. John Wiley & Sons, 1983.

[S] Snyder D. L., **随机点过程 （梁之舜，邓永录译）. 人民教育出版社，１９８１.**

[T] Tijms, H. C., Stochastic Models, An Algorithmic Approach. John Wiley & Sons, 1994.

[To]Tong, J.H., Nonlinear Time Series, A Dynamical System Approach, Oxford Univ. Press, 1990.
[W] Winkler, G., Image Analysis, Random Fields and Dynamic Monte Carlo Methods. A Mathematical Introduction. Springer, 1995.
[Wg] **王梓坤**, **随机过程论**. **科学出版社**, 1965.
[X] **谢衷洁**, **时间序列分析**. **北京大学出版社**, 1990.
[YWL] **严士健**, **王隽骧**, **刘秀芳**, **概率论基础**. **科学出版社**, 1982.
[Z] **张尧廷**, **金融市场的统计分析**. **广西师范大学出版社**, 1998.

名词索引

E

F

L

M

N

O

P

Q

R

S，(s)

T

其它

龚光鲁，钱敏平著 应用随机过程教程及其在算法与智能计算中的应用

清华大学出版社,2003

第１章　概率论精要回顾与补充

1 基本框架与典型分布

1.1　概率

定义１．１　记$\Omega=\{w: w$为一个基本事件，即随机试验的一个可能结果$\}$，它称为样本空间．设$\mathcal{F}$是Ω的某些集合(称为事件)组成的类, 它如果满足

$$\text{由 } A, A_n \in \mathcal{F}(n\geq 1) \text{ 能推出 } \bigcup_{n=1}^{\infty} A_n, \bigcap_{n=1}^{\infty} A_n, \overline{A} \overset{D}{=} W - A \in \mathcal{F}\ , \tag{1. 1}$$

就称为一个事件体(σ 代数)．如果在 $\mathcal{F}$ 上定义了一个非负函数$P(A), \forall A\in \mathcal{F}$, 满足: $P(\Omega)=1$， 而且对于任意 $A_i \in \mathcal{F}, (i\geq 1)$，只要 $A_i \cap A_i \neq \varnothing\ (i\neq j)$， 就有 $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$，则称$P(A)$为事件$A$的概率.

1. 2　随机变量

定义１．２　一个随机地取实数值的量称为随机变量, 定义随机变量x 的分布函数为 $F(x)=P(x\leq x)$．我们用$x \overset{d}{=} h$ 表示x 与h同分布.

１． 随机变量x 的数值函数$g(x)$的数学期望(均值)

定义１．3　离散随机变量x 的概率函数（概率分布）定义为

$$p(x) \overset{\Delta}{=} P(x=x) \qquad (x=x_1, x_2, ...), \qquad p(x_i)=p_i, \tag{1. 2}$$

其分布函数为 $F(x)=\sum_{x_i\leq x} p_i$, 而数值函数$g(x)$的数学期望为

$$Eg(x)=\sum_{,} g(x_i)p_i\ .$$

如果x 只取非负整数值$P(x=n)=p_n$, 则有另一个计算公式：

$$Ex=\sum_{n} P(x>n),\ \ Ex^2=\sum_{n,m\geq 0} P(x>n,m)\,. \tag{1. 3}$$

(证明: 左=$E(\sum_{n,m\geq 0} I_{\{n<x\}} I_{\{m<x\}})=\sum_{n,m\geq 0} E(I_{\{n<x\}} I_{\{m<x\}})$=右).

连续型随机变量x 的分布密度为 $p(x)$,分布函数为 $F(x)=\int_{-\infty}^{x} p(t)dt$,数值函数 $g(x)$ 的期望为

$$Eg(x)=\int_{-\infty}^{+\infty} g(x)p(x)dx .$$

如果x 只取非负值，则有另一个计算公式

$$Ex=\int_{0}^{\infty} P(x>x)dx .$$

设x 是以$a\,(0<a<1)$ 的概率取一个分布函数为$F_d(x)=\sum_{x_i\le x} p_i$ 的一个离散随机变量，而以$1-a$ 的概率取另一个分布函数为$F_c(x)=\int_{-\infty}^{x} p(t)dt$ 的一个连续型随机变量， 那么x 的分布函数就应该为 $F(x)=aF_d(x)+(1-a)F_c(x)$， 而其数值函数$g(x)$ 的数学期望为

$$Eg(x)=a\sum_{,} g(x_i)p_i+(1-a)\int_{-\infty}^{+\infty} g(x)p(x)dx \text{.}$$

对一般情形的随机变量x，设其分布函数为$F(x)$，则x 的函数 $g(x)$ 的数学期望可粗略地定义为 Stieltjes 积分

$$Eg(x)=\lim_{\substack{n\to\infty \\ \max_i (t_{i+1}^{(n)}-t_i^{(n)})\to 0}} \sum_i g(t_i^{(n)})[F(t_{i+1}^{(n)})-(F(t_i^{(n)})],$$

此极限记为 $\int_{-\infty}^{+\infty} g(x)dF(x)$， 这里$\{t_i^{(n)}\}$是[-n, n]的一个划分. 这种积分的运算规律及近似计算与普通积分类似.

我们有

$$\text{如果}\sum_k x_k<\infty,(x_k\ge 0)，\quad \text{则有}\quad E(\sum_k x_k)=\sum_k Ex_k \text{.}$$

这个等式说明了求无穷和与取期望可以交换次序的条件. 鉴于此公式很直观, 且很有用, 所以我们引述于此. 而它的证明需要用到测度论的知识，故而从略.

2.方差与矩母函数

定义１．４　随机变量x 的方差定义为

$$Varx=E(x-Ex)^2=Ex^2-(Ex)^2 .$$

矩母函数为

$$M(z)=Ee^{zx} ,$$

如果它有限， 它不仅包含了一切阶矩: $\{Ex^k$ （k 阶矩）$\}$ 的信息, 而且此时x 的分布也可由矩母函数唯一地确定. (如果矩母函数不是有限, 则人们用 “纯虚的矩母函数”， 即特征函数

$$j(t)=Ee^{itx}$$

来代替它).

1．3　d-维随机向量

d 维随机向量 $\vec{x}$ 的分布函数为 $F(\vec{x})=P(\vec{x}\le\vec{x})$，这里 $\dot{x}=\begin{pmatrix}x_1\\ \vdots\\ x_d\end{pmatrix}$， $\vec{x}\le\vec{x}$ 是指 $x_1\le x_1,...,x_d\le x_d$， 而数学期望和方差阵分别为

$$E\vec{x}=\begin{pmatrix}Ex_1\\ \vdots\\ Ex_d\end{pmatrix},\Sigma_{\bar{x}}=(Cov(x_i,x_j)_{i,j\le d}\ .$$

(x,h) 的协方差为 $Cov(x,h)=E(xh)-ExEh$，相关系数为 $r_{xh}=\dfrac{Cov(x,h)}{\sqrt{Varx}\sqrt{\operatorname{var}h}}$.

$\vec{x}$ 的矩母函数定义为

$$M(\vec{z})=Ee^{\dot{z}^T\dot{x}}, \tag{1．4}$$

其中 $\vec{z}^T=(z_1,...,z_d)$，而$(\quad)^T$ 表示转置.

$\vec{x}$ 的特征函数定义为

$$F(\dot{t})=Ee^{i\dot{t}^T\dot{x}}. \tag{1．5}$$

其中 $\dot{t}^T=(t_1,...,t_d)$

离散随机向量 $\vec{x}$ 的概率函数(概率分布)为

$$p(\dot{x})\overset{\Delta}{=}P(\dot{\xi}=\dot{x}),\qquad(\dot{x}=\dot{x}_1,\dot{x}_2,...).$$

设连续型随机向量 $\vec{x}$ 的密度为 $p(\vec{x})$， 则其分布函数为 $F(\vec{x})=\int_{-\infty}^{\bar{x}}p(\dot{t})d\dot{t}$。同样有 $Eg(\vec{x})=\int g(\vec{x})p(\vec{x})d\vec{x}$.

1．4　独立性

定义1．5　随机变量组 $\{x_1,...,x_n\}$ 称为独立，如果

$$P(x_1\le x_1,\cdots,x_n\le x_n)=P(x_1\le x_1)\cdots P(x_n\le x_n)\ .$$

随机变量组 $\{x_1,...x_n\}$ 独立 $\Leftrightarrow M(\vec{z})=M_1(z_1)\cdots M_n(z_n)$（其中 $M_i(z_i)$ 是 x_i 的矩母函数 $\Leftrightarrow \varphi(\vec{\lambda})=\varphi_1(\lambda_1)\cdots\varphi_n(\lambda_n)$ （其中 $\varphi_i(\lambda_i)$ 是 x_i 的特征函数）.

随机变量组$\{x_1,...x_n\cdots\}$称为独立，如果对任意 n，$\{x_1,...,x_n\}$都是独立的.

两个随机向量$\vec{x},\vec{h}$ 称为独立， 如果 $\forall A,B \Rightarrow \vec{x}\in A$与$\vec{h}\in B$都独立，这等价于

$$\forall f,g \Rightarrow E[f(\vec{x})g(\vec{h})] = Ef(\vec{x})Eg(\vec{h}). \qquad (1.6)$$

设随机变量x,h 独立且分别具有密度$f(x),g(x)$，则其和$x+h$具有密度(称为f,g的卷积)

$$(f*g)(x)\overset{\Delta}{=}\int f(x-u)g(u)du = (g*f)(x). \qquad (1.7)$$

特别地，如果在$x<0$时$f(x)=g(x)=0$，则上面的$f*g$ （或$g*f$ 的积分表示化为

$$(f*g)(x)\overset{\Delta}{=}\int_0^x f(x-u)g(u)du. \qquad (1.8)$$

1，5 Chebyshev 不等式

设x为随机变量，$Varx$为其方差，则有 Chebyshev 不等式 $P(|x-Ex|\ge d)\le \dfrac{Varx}{d^2}$.

Chebyshev 不等式给出了用方差来估计随机变量与它的数学期望的偏差超过某个值的概率的上界．它的优点是不依赖随机变量的分布．但是正因为如此，它就很粗糙．例如，如果已知x服从正态分布，即$x\sim N(m,s^2)$，那么下面的上界估计就更为精确:

$$P(|x-Ex|\ge d) = 2(1-\Phi(\frac{d}{s^2})),$$

其中$\Phi(x)=\int_{-\infty}^x \frac{1}{\sqrt{2p}}e^{-\frac{1}{2}z^2}dz$ 是标准正态分布$N(0,1)$的分布函数.

1，6 基本极限与基本极限定理(大数定律与中心极限定理)

定义１．６ 如果随机变量序列x_n与随机变量x间满足

$$\forall e>0 \Rightarrow P(|x_n-x|\ge e)\xrightarrow{n\to\infty} 0, \qquad (1.9)$$

则称随机变量序列x_n依概率收敛到随机变量x，记为$x_n\xrightarrow{p}x$． 此定义的含义为，如果忽略一个小的概率ε， 那么x_n可以近似x.

又若 对于任意的分量$i\le d$都有$x_i^{(n)}\xrightarrow{p}x_i$，则称为$\vec{x}^{(n)}\xrightarrow{p}\vec{x}$ （其中$\vec{x}^{(n)}=(x_1^{(n)},\cdots,x_d^{(n)})$，$\vec{x}=(x_1,\cdots,x_d)$.

对于连续函数$f(\vec{x})$，我们有

$$\vec{x}^{(n)}\xrightarrow{p}\vec{x} \quad \Rightarrow f(\vec{x}^{(n)})\xrightarrow{p} f(\vec{x}). \qquad (1.10)$$

概率论中最重要的定理之一，就是大数定律，它断定：若x_n为独立同分布的随机变量序列，且$Ex_n = m \ (n=1,2,\cdots)$，则 $\frac{x_1+\cdots+x_n}{n} \xrightarrow{p} m$.

注　若x_n为非负的独立同分布的随机变量序列，而$Ex_n = +\infty (n=1,2,\cdots)$，则 $\frac{x_1+\cdots+x_n}{n} \xrightarrow{p} +\infty$，**意即**

$$\forall C > 0 \Rightarrow P(\xi_n \le C) \xrightarrow{n\to\infty} 0 .$$

此结论可视为大数定律的推广．

定义１．７　如果随机变量序列x_n与随机变量x之间满足

$$E|x_n - x|^2 \to 0 \text{，} \tag{1. 11}$$

则称随机变量序列x_n均方收敛到随机变量x，记为$x_n \xrightarrow{L^2} x$．

由 Chebyshev **不等式立刻可以得到均方收敛一定能推出依概率收敛.**

定义１．８　如果随机变量序列h_n与随机变量h之间满足 $P(h_n \xrightarrow{n\to\infty} h) = 1$，**即"$h_n \to h$"是一个概率为** 1 **的事件，则称随机变量序列h_n概率为１收敛到随机变量h，记为** $h_n \xrightarrow{a.e.} h$．**这里** a.e. **是** almost everywhere **的缩写.**

概率为 1 **收敛一定可以推出概率收敛**

(**这个事实的证明需要用到一点测度论或实变函数的知识．其证明如下：事件$\{h_n \to h\}$就是"任给$\frac{1}{m} > 0$，必存在n_0只要$n \ge n_0$，就有$|h_n - h| < \frac{1}{m}$"．把它写成式子，就是** $\bigcap_m \bigcup_{n_0=1}^{\infty} \bigcap_{n=n_0}^{\infty} \{|h_n \to h| < \frac{1}{m}\}$．**故由$P(h_n \xrightarrow{n\to\infty} h) = 1$推出**

$$P(\bigcup_{n_0=1}^{\infty} \bigcap_{n=n_0}^{\infty} \{|h_n \to h| < \frac{1}{m}\}) \ge P(\bigcap_m \bigcup_{n_0=1}^{\infty} \bigcap_{n=n_0}^{\infty} \{|h_n \to h| < \frac{1}{m}\}) = 1 .$$

由此即能推出 $\lim_{n\to\infty} P(|h_n - h| < \frac{1}{m}) = 1$).

若随机变量序列x_n独立同分布且期望有限，则 $\frac{x_1+\cdots+x_n}{n} \xrightarrow{a.e.} Ex_1 (n \to \infty)$**，此结论称为强大数定律．．**

这个定理在非数学专业的概率论课程中，一般较少论及，主要因为"随机变量列的收

敛”是一个什么事件不易说清楚。再则，其证明也较为复杂．然而这个定理的概率直观内容是非常清楚的，因此我们在此特别列出.

注　我们同样有：若随机变量序列 ξ_n **为非负的，并且独立同分布，而** $E\xi_n=+\infty$**，则**

$$\frac{\xi_1+\cdots+\xi_n}{n}\xrightarrow{a.e.}+\infty .$$

定义１．８　设ξ **为连续型随机变量,其分布函数为** F(x)，**如果随机变脸序列**ξ_n **的分布函数收敛到** F(x)**，则称**ξ_n **依分布收敛到**ξ，**记为**$\xi_n\xrightarrow{d}\xi$ **.**

依分布收敛可以广到一般的非连续型的随机变量ξ **。　即：如果在随机变量**ξ **的分布函数** $F(x)$ **的所有的连续点** x **上有：随机变量序列**ξ_n **的分布函数收敛到** F(x)，**则也称为**ξ_n **依分布收敛到**ξ．**仍记为** $\xi_n\xrightarrow{d}\xi$ （**例如,**ξ **遵从** Poisson **分布**（**参见后面的典型分布**—1．**7段**）**就是一种常见的情形**）.

依分布收敛也称为弱收敛，也记为 $\xi_n\xrightarrow{w}\xi$．**在概率论的理论中已经证明了**：

$$\xi_n\xrightarrow{d}\xi \Leftrightarrow Ee^{il\xi_n}\to Ee^{il\xi}$$

$$\Leftrightarrow Ef(\xi_n)\to Ef(\xi) \quad \text{（f 为任意有界连续函数）.} \qquad (1.12)$$

从概率收敛一定能推出依分布收敛.（**证明：记**ξ_n,ξ **的分布函数为** F_n,F .**固定** x，**对于** $x'<x$ **有** $\{\xi\le x'\}\subset\{\xi_n\le x\}+\{\xi_n>x,\xi\le x'\}$．**于是，由上式和** $\xi_n\xrightarrow{p}\xi$ **推出** $F(x')\le F_n(x)+P(\xi_n>x,\xi\le x')\le F_n(x)+P(|\xi_n-\xi|\ge x-x')$．**因此** $F(x')\le\lim_m\inf_{n\ge m}F_n(x)$．**对称地利用** $x''>x$，**可以证明** $F(x'')\ge\lim_m\sup_{n\ge m}F_n(x)$．**连起来成为** $F(x')\le\lim_m\inf_{n\ge m}F_n(x)\le\lim_m\sup_{n\ge m}F_n(x)\le F(x'')$．**如果** F **在** x **连续，那么令** $x',x''\to x$，**便得** $F(x)\le\lim_m\inf_{n\ge m}F_n(x)\le\lim_m\sup_{n\ge m}F_n(x)\le F(x)$．**这就是** $F_n(x)\to F(x)$）.

反之，从依分布收敛却不能推出概率收敛，除非极限随机变量为一个常数.
一般地，我们有

$$\xi_n\xrightarrow{p}\xi,\eta_n\xrightarrow{d}\eta \quad\Rightarrow\quad \xi_n+\eta_n\xrightarrow{d}\xi+\eta . \qquad (1.13)$$

但是从$\xi_n\xrightarrow{d}\xi,\eta_n\xrightarrow{d}\eta$ **并不能推出** $\xi_n+\eta_n\xrightarrow{d}\xi+\eta$ **.**

依分布收敛也可以用于近似计算概率.

对于随机向量, 类似地也有依分布收敛及收敛的充要条件. 于是有

$$(x_1^{(n)},\cdots,x_d^{(n)})\xrightarrow{d}(x_1,\cdots,x_d) \quad\Leftrightarrow\quad \forall(s_1,\cdots,s_d),\sum_{i=1}^{d}s_ix_i^{(n)}\xrightarrow{d}\sum_{i=1}^{d}s_ix_i . \ (1.14)$$

概率论中另一个最重要的定理是中心极限定理, 其叙述如下

若随机变量序列 x_n 为独立同分布, $Ex_n=m$ 且 $Varx_n=s^2(n=1,2,\cdots)$, 则

$$\frac{x_1+\cdots+x_n-nm}{\sqrt{n}s}\xrightarrow{d}N(0,1).$$

中心极限定理对 d 维随机变量也是成立的.

定理1.9 (Polya 定理) 设随机变量序列 x_n 和随机变量 x 满足 $x_n\xrightarrow{d}x$, 且 x 的分布函数 $F_x(x)$ 是连续函数, 则 x_n 的分布函数 $F_{x_n}(x)$ 一致收敛到 x 的分布函数 $F_x(x)$.

此结论称为 Polya 定理, 其证明是数学分析的一个习题.

推论1.10 若随机变量序列 x_n 为独立同分布, $Ex_n=m$ 且 $Varx_n=s^2(n=1,2,\cdots)$, 则

$$Sup_{-\infty<x<\infty}\left|P(\xi_1+\cdots+\xi_n\le x)-\frac{1}{\sigma\sqrt{2\pi\cdot n}}\int_{-\infty}^{x}e^{-\frac{(u-n\mu)^2}{2n\sigma^2}}du\right|\xrightarrow{n\to\infty}0.$$

这可以理解为 $x_1+\cdots+x_n$ 有近似分布 $N(E(x_1+\cdots+x_n),Var(x_1+\cdots+x_n))$.

可以用特征函数或矩母函数来表示依分布收敛, 具体结论如下

若随机变量序列 x_n 的特性函数 $j_n(l)\xrightarrow{n\to\infty}j(l)$, 且 $j(l)$ 在 $l=0$ 点连续, 则 $\varphi(\lambda)$ 必是某个随机变量 x 的特征函数, 而且 $x_n\xrightarrow{d}x$.

注 1 以上结论比(1.12)式要好用. 因为在(1.12)式中需要事先知道极限函数确实是特征函数, 而此需要正是在应用时难以判别的. 而此断言正好给出了判断一个函数为特征函数的充分条件, 所以这是非常有用的.

注 2 此结论在内容上易于理解, 在应用上也简单方便, 但是其证明则需要用到超出本书范围的 Fourier-Stieltjes 分析这个数学工具. 而在一般初等概率论中也并不给出此结论的推导. 有兴趣的读者可以选阅较高一些水平的概率论教科书, 例如, 严士健等人编写的《概率论》.

注 3 对于 d 维情形, 相应的结论仍然正确.

1.7 典型分布

离散随机变量的典型分布有

[Bernoulli(二项)分布 $B(N,p)$] 概率函数为

$$p(x)=\binom{N}{x}p^x(1-p)^{N-x}\ (0<p<1,x=0,1,\ldots,N),$$ 数学期望为 Np,

方差为 $Np(1-p)$**，矩母函数为** $M(z)=(pe^{z}+(1-p))^{N}$**.**

[**Poisson 分布** $Poisson_{l}$]　**概率函数为** $p(x)=e^{-l}\frac{l^{x}}{x!}$　$(l>0, x=0,1,2,...)$**，**

数学期望为 l**，　方差为** l**，　矩母函数为** $M(z)=e^{l(e^{z}-1)}$**.**

[**几何分布**]　**概率函数为** $p(x)=p(1-p)^{x-1}$　$(x=1,2,...)$**，**

数学期望为 $\frac{1-p}{p}$**，　方差为** $\frac{1-p}{p^{2}}$**，矩母函数为** $M(z)=\frac{pe^{z}}{1-(1-p)e^{z}}$**.**

[**负二项分布 (Pascal 分布)**]　r **个具有独立同几何分布的随机变量的和的分布称为负二项分布．概率函数为**　$p(x)=C_{r-1}^{r-x}p^{r}(1-p)^{x-r}$　$(x=r,r+1,...)$**，**

数学期望为 $\frac{r(1-p)}{p}$**，　方差为** $\frac{r(1-p)}{p^{2}}$**，矩母函数为** $M(z)=(\frac{pe^{z}}{1-(1-p)e^{z}})^{r}$**.**

[**多项分布**]　**概率函数为** $p(\vec{x})=C_{n}^{n_1,\cdots,n_k}p_1^{n_1}\cdots p_k^{n_k}$　$(\vec{x}=(n_1,\cdots,n_k), n_1+\cdots+n_k=n)$,

$(p_1,\cdots,p_k>0, p_1+\cdots+p_k=1)$, **其中** $C_{n}^{n_1,\cdots,n_k}=\frac{n!}{n_1!\cdots n_k!}$.

连续型随机变量的典型分布

[**正态分布** $N(m,s^{2})$]　**分布密度为** $p(x)=\frac{1}{\sqrt{2p}s}e^{-\frac{(x-m)^{2}}{2s^{2}}}$　$(s>0)$**，**

数学期望为 m**，　方差为** s^{2}**，矩母函数为** $M(z)=e^{mz+\frac{1}{2}s^{2}z^{2}}$，**特征函数为** $j(t)=e^{imt-\frac{1}{2}s^{2}t^{2}}$**.**

[**指数分布** Exp_{l}]　**分布密度为** $p(x)=le^{-lx}I_{[0,\infty)}(x)$　$(l>0)$**，**

数学期望为 $\frac{1}{l}$**，　方差为** $\frac{1}{l^{2}}$**，矩母函数为** $M(z)=\frac{\lambda}{\lambda-z}$

指数分布是唯一的一个取值于$[0,\infty)$**的无记忆分布，即满足：对于任意**$s,t>0$，**恒有**

$$P(x>s+t\,|\,x>t)=P(x>s).$$

[**均匀分布** $U[a,b]$]　**分布密度为**　$p(x)=\frac{1}{b-a}I_{[a,b]}(x)$**，**

数学期望为 $\frac{a+b}{2}$**，　方差为** $\frac{(b-a)^{2}}{12}$**，矩母函数为** $M(z)=\frac{e^{bz}-e^{az}}{(b-a)z}$.

[Gamma 分布 （$\Gamma(a,l)$分布）] 分布密度为$p(x)=\frac{\lambda^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}I_{[0,\infty)}(x)\quad(a,l>0)$,

数学期望为 $\frac{a}{l}$, **方差为** $\frac{a}{l^2}$, $Ex^k=\frac{\Gamma(a+k)}{l^k\Gamma(a)}$, **矩母函数为** $M(z)=(\frac{\lambda}{\lambda-z})^{\alpha}$.

（**注**: $\Gamma(a)=\int_0^{\infty}t^{a-1}e^{-t}dt$, $\Gamma(x+1)=x\Gamma(x)$, $\Gamma(n+1)=n!$).

[逆 Gamma 分布（$IG(a,l)$分布］ 设$x\sim G(a,l)$，则$h=\frac{1}{x}$的分布称为逆 Gamma 分布.

（逆 Gamma 分布常常在方差的 Bayes 统计中，用作方差的先验分布）

[Erlang 分布（记为$Erlang_{n,\lambda}$)] 它是n个独立的 Exp_l 随机变量的和的分布．它就是 Γ（n,l)分布.

[$c^2(n)$ 分布］ 它是n个独立的 $N(0,1)$ 随机变量的平方和的分布．分布密度为

$p(x)=\frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}x^{\frac{n}{2}-1}e^{-\frac{x}{2}}I_{[0,\infty)}(x)\quad(l>0)$， **数学期望为** n， **方差为** $2n$.

[Beta 分布（$B(a,b)$分布)]

分布密度为 $p(x)=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}x^{a-1}(1-x)^{b-1}I_{[0,1]}(x)\quad(a,b>0)$， **数学期望为** $\frac{a}{a+b}$,

方差为 $\frac{ab}{(a+b)^2(a+b+1)}$，$k$ **阶矩为** $Ex^k=\frac{a(a+1)\cdots(a+k-1)}{(a+b)(a+b+1)\cdots(a+b+k-1)}$.

[指数族分布］ (是包含上述多种分布的概括与推广） 分布密度为 $p(x)=C(\vartheta)h(x)e^{\sum_{k=1}^{m}\varphi_k(\vartheta)T_k(x)}$.

[Weibull 分布 $W(a,l)$] **分布密度为** $p(t)=l\cdot a\cdot t^{a-1}e^{-l\cdot t^{a}}I_{[0,\infty)}(t)(a,l>0)$，

数学期望为 $l^{-\frac{1}{a}}\Gamma(1+\frac{1}{a})$, **方差为** $l^{-\frac{2}{a}}[\Gamma(1+\frac{2}{a})-(\Gamma(1+\frac{1}{a}))^2]$.

（若$x\sim\exp_l$， 则$h=x^{\frac{1}{a}}\sim W(a,l)$）.

[广义 Gamma 分布] **分布密度为** $p(t)=\frac{b}{\Gamma(k)}s^{-bk}t^{-bk-1}e^{-(\frac{t}{s})b}I_{[0,\infty)}(t)$.

[**截尾正态分布**] **分布函数为** $F(t)=C(\Phi(t)-\frac{1}{2})I_{[0,\infty)}(t)$，**其中** $\Phi(x)$ **为** $N(0,1)$ **的分布函数**, C **为规格化常数**.

[**Pareto 分布**] **分布密度为** $p(x)=ra^r\frac{1}{x^{r+1}}I_{[a,\infty)}(x)$，

数学期望为 $\frac{ra}{r-1},(r>1)$， **方差为** $\frac{ra^2}{(r-1)(r-2)},(r>2)$.

另外，分布密度为 $p(x)=\frac{\alpha\lambda^{\alpha}}{(\lambda+x)^{\alpha+1}}I_{(-\lambda,0]}(x)$ **的分布也称为** Pareto **分布**. **它出现在经济学中**, **例如**, **在成熟的市场经济社会中**, **财富的占有人数的分配比例近似地呈现为**Pareto**分布**. **若**$x\sim Pareto(r,a)$, **则**$h=\frac{r}{l}\log\frac{x}{a}\sim Exp_l$.

[**极值分布** $E(a,b)$，**也称为**$Gumble(a,b)$ **分布**]

分布函数为$F(x)=(1-e^{-\exp(-\frac{x-a}{b})})I_{[0,\infty)}(x)$, **数学期望为** $a+Cb$ （C **为** Euler **常数**), **方差为** $\frac{p^2b^2}{6}$, **特征函数为**$j(t)=e^{iat}G(1-ib\cdot t)$.

(**若**$x\sim Exp_1$，**则**$a-b\log x\sim Gumble(a,b)$).

[**Logistic**(a,b) **分布**] **分布函数为** $F(x)=\frac{1}{1+e^{-\frac{x-a}{b}}}$,

数学期望为 a, **方差为** $\frac{p^2b^2}{3}$, **特征函数为** $j(t)=e^{iat}\frac{ip\cdot b\cdot t}{\sin(ip\cdot b\cdot t)}$.

[**逆 Gauss 分布**] **分布密度为** $p(x)=\sqrt{\frac{l}{2p\cdot x^3}}e^{-\frac{l(x-a)^2}{2a^2x}}I_{[0,\infty)}(x)(l,a>0)$.

数学期望为 a， **方差为** $\frac{a^3}{l}$.

[**复合 Poisson 分布**] **设**$x_1,x_2,\ldots$**为独立同分布**, **其分布函数均为**$F(x)$, **为简单起见,我们假定它具有密度函数** $f(x)$, N **是一个与**$\{x_1,x_2,\ldots\}$**独立的随机变量**, **且遵从** $Poisson_l$, **则** $h=x_1+x_2+\cdots+x_{\mathrm{N}}$ **的分布称为复合** Poisson **分布**, **其分布密度函数为** $f_h(x)\overset{\Delta}{=}\sum_{k=0}^{\infty}e^{-l}\frac{l^k}{k!}f^{*k}(x)$, **其中** $f^{*k}(x)$ **为** $f(x)$ **的**k **次卷积** (**参见** (1. 7)**式**).

复合 Poisson 分布的数学期望为 $E\eta = \lambda \cdot E\xi_1$，方差为 $Var(\eta) = \lambda \cdot E{\xi_1}^2$，矩母函数为

$M(z) = e^{l(f(z)-1)}$ ， 其中 $f(z)$ 为 x_1 的矩母函数 (其证明则需要用后面 2. 4 段的 Wald 等式).

[对数正态分布] (即 $\ln x \sim N(m, s^2)$)] 分布密度为 $p(x) = \dfrac{1}{\sqrt{2p}\, s x} e^{-\frac{(\log x - m)^2}{2s^2}}$，

数学期望 $e^{m+\frac{1}{2}s^2}$， 方差为 $e^{2m+s^2}(e^{s^2}-1)$.

[Cauchy 分布] 分布密度为 $p(x) = \dfrac{c}{p[(x-g)^2 + c^2]}$,

数学期望不存在， 特征函数为 $j(t) = e^{igt - c|t|}$.

典型的多维分布

[d-维正态分布 $N(\vec{m}, \Sigma)$] 分布密度为 $p(\vec{x}) = \dfrac{1}{(2p)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(\vec{x}-\vec{m})^T \Sigma^{-1} (\vec{x}-\vec{m})}$，

其中 $\Sigma = \left(s_{ij}\right)_{i,j \le d}$ 为对称正定矩阵.

矩母函数为 $M(\vec{z}) = e^{\vec{\mu}^T \vec{z} + \frac{1}{2}\vec{z}^T \Sigma \vec{z}}$，特征函数为 $\varphi(\vec{\lambda}) = e^{\vec{\mu}^T \vec{\lambda} + \frac{1}{2}\vec{\lambda}^T \Sigma \vec{\lambda}}$.

[多维 Beta 分布 (Dirichlet$(a_1, \cdots, a_k)$ 分布)]

分布密度为 $p(\vec{x}) = \dfrac{\Gamma(a_1)\cdots\Gamma(a_k)}{\Gamma(a_1 + \cdots + a_k)} {x_1}^{a_1 - 1} \cdots {x_k}^{a_k - 1} I_{\{x_1 + \cdots + x_k = 1, x_1, \cdots, x_k \ge 0\}}(x_1, \cdots x_k)$.

[d-维对数正态分布（$\ln \vec{x}$ 的分布，其中 $\vec{x} \sim N(\vec{m}, \Sigma)$）]

分布密度为 $p(\vec{x}) = \dfrac{1}{(2p)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}\, x_1 \cdots x_d} e^{-\frac{1}{2}(\ln \vec{x} - \vec{m})^T \Sigma^{-1} (\ln \vec{x} - \vec{m})}$，

$\Sigma = \left(s_{ij}\right)_{i,j \le d}$ 为对称正定矩阵， $\ln \vec{x} = (\ln x_1, \ldots, \ln x_d)^T$ ，相关系数为

$r_{ij} = \dfrac{e^{s_{ij}} - 1}{(e^{s_{ii}} - 1)(e^{s_{jj}} - 1)}$.

[d-维 Gauss 分布] 它是 d-维正态分布的推广，包含 d-维正态分布与 “ 退化的 d-维正态分布 ”两类，后者没有密度且不易写出它的分布函数，因此对这种随机向量 $\vec{x}$ 的分布用它的特征函数描述更为方便，它的特征函数在形式上与正态分布的特征函数是一样的， 仍为 $\varphi(\vec{\lambda}) = e^{\vec{\mu}^T \vec{\lambda} + \frac{1}{2}\vec{\lambda}^T \Sigma \vec{\lambda}}$， 所不同的是： 这里的 Σ 为对称非负定矩阵，即可以退化 (由于这

里的 $\varphi(\lambda)$ 是 d-维正态分布特征函数 $\varphi_n(\vec{z}) \overset{\Delta}{=} e^{\vec{\mu}^T \vec{\lambda} + \frac{1}{2}\vec{\lambda}^T(\Sigma + \frac{1}{n}I)\vec{\lambda}}$ 的极限，而第 1．6 段用特征函数表示依分布收敛的结论正说明了它确是特征函数)．此时随机向量可以只分布在较低维的空间上, 甚至可以为常数), 这个分布仍记为 $N(\vec{m}, \Sigma)$．

注 1 由 Gauss 分布的定义可以看出:

$$\vec{x} \sim N(\vec{m}, \Sigma) \Rightarrow A\vec{x} + \vec{b} \sim N(A\vec{m} + \vec{b}, A\Sigma A^T).$$

注 2 随机向量 $\vec{x}$ 遵从 Gauss 分布等价于: 对于任意 d 维的向量 $\vec{a}$, $\vec{a}^T\vec{\xi}$ 都遵从 Gauss 分布（即一维正态分布或常数）.

Gauss 分布有一个重要的特性, 就是关于依分布收敛的封闭性, 即: 若 $x_n \sim N(m_n, s_n^2)$, 且 $x_n \xrightarrow{d} x$, 则 $\exists m, s^2$, 使 $m_n \to m, s_n^2 \to s^2$, 且 $x \sim N(m, s^2)$.

(这一性质的证明需要用一点复变函数论知识, 在此略去).

[多维指数族分布] 若二维随向量 (x, h) 满足:

$$P(x > t, h > s) = \exp\{-l_1 t - l_2 s - l_3 (t \vee s)\},$$

则称为服从二维 Poisson 分布.

[d 维 Cauchy 分布] 分布密度为 $p(\vec{x}) = \Gamma(\frac{d+1}{2}) \dfrac{c}{[p(|\vec{x} - \vec{g}|^2 + c^2)]^{\frac{d+1}{2}}}$,

特征函数为 $j(\vec{t}) = e^{i\vec{g}^T\vec{t} - c|\vec{t}|}$.

1.8 次序随机变量的分布

设 $\xi_1 = \xi_1(\omega), \cdots, \xi_n = \xi_n(\omega)$ 为独立同分布的随机变量, 且具有分布密度 $p(x)$. 当 ω 固定时（即取定一个基本事件 ω 时）$\xi_1(\omega), \cdots, \xi_n(\omega)$ 按次序重新排列记为

$$\xi_1^*(\omega) \le \cdots \le \xi_n^*(\omega),$$

称为次序统计量. 那么 $(\xi_1^*, \cdots, \xi_n^*)$ 的联合分布密度为

$$g(x_1, \cdots, x_n) = n!\, p(x_1) \cdots p(x_n) I_{x_1 < \cdots < x_n},$$

即

$$g(x_1, \cdots, x_n) = \begin{cases} n!\, p(x_1) \cdots p(x_n), & x_1 < \cdots < x_n, \\ 0 & \text{其他情形}. \end{cases}$$

2 条件概率，条件分布，条件(数学)期望.

2.1 条件概率

设 A 是具有正概率的事件，对于任意事件 B，称 $P(B\mid A)=\dfrac{P(A\cap B)}{P(A)}$ 为在已知事件 A 发生的条件下，事件 B 发生的条件概率．简称事件 B 对于事件 A 的条件概率．

全概率公式　若正概率事件序列 $\{A_n\}_{n\geq 1}$ 是样本空间 Ω 的一个划分，即:

$$\forall A_n \in \mathrm{F},\ (n\geq 1),\ \bigcup_{n\geq 1} A_n=\Omega,\qquad A_n\cap A_m\neq\varnothing\qquad (n\neq m)$$

那么,对任意 $B\in$ F，有

$$P(B)=\sum_{n\geq 1}P(B\mid A_n)P(A_n) \tag{1. 15}$$

Bayes 公式

$$P(A_i\mid B)=\frac{P(B\mid A_i)P(A_i)}{P(B)}\text{.}$$

Bayes **公式是常与全概率公式结合起来使用的．**

2．2　**条件分布**

设二维随机向量 (x,h) 能取的值为 $(x_i,y_j)(i,j=1,2,\cdots)$．在离散随机变量 h 取值 y 的条件下，离散随机变量 x 的条件分布是指下面的分布表

$$\begin{pmatrix}\cdots & x_i & \cdots\\ \cdots & P(x=x_i\mid h=y) & \cdots\end{pmatrix},$$

其中

$$P(x=x_i\mid h=y)=\begin{cases}P(x=x_i\mid h=y_j) & (y=y_j)\\ 0 & (y\neq y_1,\cdots,y_j,\cdots)\end{cases}.$$

连续型随机变量的条件分布密度

定义１．１１　设随机向量 $(x,h)\sim$ 密度 $p(x,y)$，则在 $h=y$ 的条件下，随机变量 x 的条件密度，记为 $p_x(x\mid h=y)$ 或 $p_{x|h}(x\mid y)$，可以通过下式定义

$$\int_{-\infty}^{x} p_x(x'\mid h=y)dx' \overset{\Delta}{=} \lim_{e\to 0} P(x\leq x\mid y-e<h<y+e)\text{.}$$

即

$$p_{x|h}(x\mid y)\overset{\Delta}{=}p_x(x\mid h=y)=\frac{p(x,y)}{p_h(y)},$$

其中 $p_h(y)=\int p(x,y)dx$ **是 h 的(边缘)分布.**

推论１．１２　（积分形式的全概率公式）

设随机向量$(x,h)\sim$ **密度**$p(x,y)$, L **为实数集合，则**

$$P(x\in L)=\int_L\int p_x(x\,|\,h=y)p_h(y)dydx \text{.}$$

事实上，我们有更为一般的全概率公式，即对于任意随机事件 A **有**

$$P(A)=\int P(A\,|\,h=y)p_h(y)dy \text{，}$$

它是后面将叙述的全期望公式的特殊情形．

推论１．１３　（积分形式的 Bayes 公式）

设随机向量$(x,h)\sim$ **密度**$p(x,y)$, **则**

$$p_h(y\,|\,x=x)=\frac{p_x(x\,|\,h=y)p_h(y)}{\int p_x(x\,|\,h=z)p_h(z)dz} \text{.}$$

多维随机变量的情形条件分布

设随机向量 $(\vec{x},,\vec{h})\sim$ **密度**$p(\vec{x},\vec{y})$, **则定义** $p_{\vec{x}}(\vec{x}\,|\,\vec{y})\overset{\Delta}{=}p_{\vec{x}}(\vec{x}\,|\,\vec{h}=\vec{y})=\dfrac{p(\vec{x},\vec{y})}{p_{\vec{h}}(\vec{y})}$ ，**其中**

$p_{\vec{h}}(\vec{y})=\int p(\vec{x},\vec{y})d\vec{x}$ **是**$\vec{h}$ **的(边缘)分布**.

例１．１４　$\begin{pmatrix}\vec{x}\\ \vec{h}\end{pmatrix}\sim N\left(\begin{pmatrix}\vec{m}_1\\ \vec{m}_2\end{pmatrix},\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}\right)$ **且**$\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}$**，**$\Sigma_{22}$ **都可逆**，**则**

$$p_{\vec{x}}(\vec{x}\,|\,\vec{y})=N\left(\overrightarrow{m_1}+\Sigma_{12}\Sigma_{22}{}^{-1}(\vec{y}-\overrightarrow{m_2}),\Sigma_{11}-\Sigma_{12}\Sigma_{22}{}^{-1}\Sigma_{21}\right)\text{.}$$

混合型的随机变量的条件分布的例子:

例１.１５　设随机向量(x,h) **的分布函数为**

$$F(x,y)=a\int_{-\infty}^{x}\int_{-\infty}^{y}p(u,v)dudv+(1-a)\sum_{\substack{x_i\le x\\ y_j\le y}}p_{ij}\qquad(0<a<1),$$

其中 $p(x,y)$**是一个分布密度**，**而**$\{p_{ij}\}$ **是一个分布律**.**那么**

$$P(x\le x\,|\,h=y)\overset{\Delta}{=}\lim_{e\to 0}P(x\le x\,|\,y-e<h<y+e)$$
$$=\begin{cases}\dfrac{\sum\limits_{x_i\le x}p_{ij}}{\sum\limits_{i}p_{ij}} & (y=y_j)\\[2ex] \dfrac{\int_{-\infty}^{x}p(u,y)du}{\int p(u,y)du} & (y\notin\{y_j:j\ge 1\})\end{cases}$$

例１.１６　设h **为取值于**$\{y_j,j\ge 1\}$ **的离散型随机变量**，**而随机向量**(x,h) **的联合分**

布可以表达为 $P(x \le x, h = y_j) = \int_{-\infty}^{x} f_j(u)du$，则$x$ 为连续型随机变量， 而且有

$$P(x \le x \mid h = y) \overset{\Delta}{=} \lim_{e \to 0} P(x \le x \mid y - e < h < y + e)$$

$$= \sum_j \frac{\int_{-\infty}^{x} f_j(u)du}{\int f_j(u)du} I_{\{y_j\}}(y).$$

2．3 条件(数学)期望

在(x,h)为连续型的或离散型的随机变量的情形时x 关于h 的条件期望的含义

对于连续型与离散型的二维随机变量这两种简单的情形，在$h = y$条件下，x 的条件期望，自然就应是：在$h = y$ 条件下x 的条件分布的数学期望．即

(1) 若 (x,h)为离散型的随机向量，则

$$E(x \mid h = y) = \sum_x x P(x = x \mid h = y\}\ ;$$

(2) 若 (x,h)为连续型的随机向量，则

$$E(x \mid h = y) = \int x p_x(x \mid y) dy = \frac{\int x p(x,y) dx}{\int p(x,y) dx}\ .$$

从这两个例子我们可以看出，在条件$h = y$ 下的条件期望$E(x \mid h = y)$是 y 的一个函数．如果我们把它记为$j(y)$，那么在条件$h = z$ 下的条件期望$E(x \mid h = z)$就是$j(z)$．这样我们就可以用随机变量$j(h)$，来表达在h 取不同z 值时的条件期望： 对任意z，当$h = z$ 时，$E(x \mid h = z) = j(z)$．可见，$j(h)$作为随机变量h 的函数，表达了x 关于h 取不同值下的条件期望．于是我们可以合理地把它记成为$E(x \mid h)$．也就是令 $E(x \mid h) \overset{\Delta}{=} j(h)$．

这样定义的$E(x \mid h)$是一个随机变量， 而且它是h 的函数， 当$h = y$时它的值为 $E(x \mid h = y)$．这个$E(x \mid h)$就称为x 关于h 的条件期望．这就比 "在$h = y$ 条件下" x 的条件期望， 在概念上大大的深入了．

需要特别强调的是，$E(x \mid h)$是随机变量， 只有当已知h 取定值的条件下， 它才有一个确定的数值．这是与普通的数学期望根本不同之处．

例１．１５（续） 在例１．１５中， 我们有

$$E(\xi \mid \eta)=\frac{\sum_i x_i p_{i\eta}}{\sum_i p_{i\eta}} I_{\{y_i\}}(\eta)+\frac{\int xp(x,\eta)dx}{\int p(x,\eta)dx} I_{R^1 \setminus \bigcup_i \{y_i\}}(\eta) .$$

例１．１６（续）　在例１.１６中，我们有

$$E(\xi \mid \eta)=\sum_j \frac{\int uf_j(u)du}{\int f_j(u)du} I_{\{y_j\}}(\eta) .$$

例１．１４（续） 设$\begin{pmatrix} \vec{x} \\ \vec{h} \end{pmatrix} \sim N\left(\begin{pmatrix} \vec{m_1} \\ \vec{m_2} \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right)$ **，而且**$\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$**与**$\Sigma_{22}$**都可逆**,

则 $E(\vec{x} \mid \vec{h}=\vec{y})$ **为分布** $p_{\vec{x}}(\vec{x} \mid \vec{y})=N\left(\overrightarrow{m_1}+\Sigma_{12}\Sigma_{22}{}^{-1}(\vec{y}-\overrightarrow{m_2}), \Sigma_{11}-\Sigma_{12}\Sigma_{22}{}^{-1}\Sigma_{21}\right)$ **的期望**,

即

$$E(\vec{x} \mid \vec{h}=\vec{y}) = \overrightarrow{m_1}+\Sigma_{12}\Sigma_{22}{}^{-1}(\vec{y}-\overrightarrow{m_2}) .$$

此时 $E(\vec{x} \mid \vec{h})$ **是** $\vec{h}$ **的线性函数**:

$$E(\vec{x} \mid \vec{h})=\overrightarrow{m_1}+\Sigma_{12}\Sigma_{22}{}^{-1}(\vec{h}-\overrightarrow{m_2}) \tag{1.16}$$

在随机向量 (x,h) **为连续型或离散型的情形时** x **关于** h **的条件期望的性质：**

在以上各种情形中所定义的随机变量 x **关于随机变量** h **的条件期望** $E(x \mid h)$ **都满足以下的基本性质**:

(E.1) 随机变量 x **关于随机变量** h **的条件期望** $E(x \mid h)$ **是** h **的函数，**

(E.2) 对任意实数集合 Λ**，有** $E[I_\Lambda(h)E(x \mid h)]=E(x \cdot I_\Lambda(h))$ **.** (1. 17)

(**我们在随机向量** $(x,h) \sim$ **密度** $p(x,y)$ **的情形验证性质**(E.1)**与**(E.2)**如下：按前面的定义我们有** $E(x \mid h)=\frac{\int xp(x,h)dx}{\int p(x,h)dx}$ **.** **它显然满足**(E.1). **而**(E.2)**的左边就是**

$$\int I_\Lambda(y)E(x \mid h=y)p_h(x,y)dy=\int I_\Lambda(y)\frac{\int xp(x,y)dx}{\int p(x,y)dx}(\int p(x,y)dx)dy .$$

$$=\int I_\Lambda(y)xp(x,y)dxdy=(E.2)\text{右边}$$

读者可自行验证，在 (x,h) **为离散随机向量或以上的混合情形时，** (E.1)**和**(E.2)**也都正确**).

可以证明由(E.1)**及**(E.2)**确定的随机变量，在不计零概率事件的差别下，是唯一的**.

上面讨论的是在 (x,h) **为连续型的或离散型的情形时的性质**. **对于最一般的任意随机变**

量(x,h)， 在理论概率论中, 人们证明了：只要随机变量x 的数学期望存在，那么就一定存在随机变量h 的一个函数$y(h)$满足： 对任意的实数集合Λ，恒有

$$E(I_\Lambda(h)y(h)) = E(x \cdot I_\Lambda(h)).$$

定义１．１７ 以上的$y(h)$就定义为随机变量x 关于随机变量h 的条件期望$E(x\mid h)$,

显见，条件期望$E(x\mid h)$是一个随机变量, 其直观含义为： 把随机变量h 看成固定时，在顾及随机变量x 与随机变量h 间的联系时x 的数学期望. 这样定义的条件期望$E(x\mid h)$当然满足(E.1)及(E.2), 其中(E.1)与(E.2)的含义为：$E(x\mid h)$是这样一个随机量, 它在h 取值已知时是完全确定的，而且在有关h 的任意一个随机事件$\{h\in\Lambda\}$上看， 它的平均与x 的平均是相等的.

例１．１８ 如果随机向量(x,h)的联合分布函数$F(x,y)$没有密度, 但是随机变量h 的边缘分布函数$F(\infty,y)$有密度 $f_2(y)$，而且对任意固定的x 而言，$\dfrac{\partial F(x,y)}{\partial y}$对$y$ 分段连续. 那么, 随机变量x 关于随机变量h 的条件分布函数（理解为条件期望$E(I_{(-\infty,x]}(x)\mid h)$）为

$$P(x\le x\mid h)=\frac{\dfrac{\partial F(x,h)}{\partial y}}{f_2(h)}.$$

（证明: 对于任意实数集合Λ，我们有

$$E[\frac{\dfrac{\partial F(x,h)}{\partial y}}{f_2(h)}I_\Lambda(h)] = \int_\Lambda \frac{\dfrac{\partial F(x,y)}{\partial y}}{f_2(y)} f_2(y)dy = \int_\Lambda \frac{\partial F(x,y)}{\partial y}dy$$

$$= \iint I_{(-\infty,x]}(u)I_\Lambda(y)F(du,dy) = E[I_{(-\infty,x]}(x)I_\Lambda(h)]\ ）.\qquad ?$$

在(E.1) 中取$\Lambda=(-\infty,\infty)$ 就得到

(E.3) $E(E(\xi\mid\eta)) = E\xi$. (1. 18)

这个公式通常称为全期望公式. 最常见的情形是，随机变量h 具有密度函数$g(y)$的情形, 此时

$$Ex\,(=E(E(x\mid h))) = \int E(x\mid h=y)g(y)dy. \qquad (1.18)'$$

它是全概率公式的推广，因为对于任意随机事件 A **，若取随机变量** $x = I_A(w)$ **时（即**

$x=\begin{cases}1 & \text{若随机事件}A\text{发生了}\\ 0 & \text{若随机事件未}A\text{发生}\end{cases}$ **），它就是全概率公式的如下的推广形式：**

$$P(A)(=E(E(I_A\,|\,h)))=\int P(A\,|\,h=y)g(y)dy\,. \qquad (1.18)''$$

无论在理论上或者在实用中，公式(1. 18)’ **与** (1. 18)” **都是十分有用的**.

用理论概率论的逼近方法还可得到：对于任意实函数$h(x)$ **恒有**

(E.4) $E(h(h)E(x\,|\,h)) = E(h(h)x)\,.$ (1. 19)

(**而**(E.2)**是它的特殊情形，即**$h(y)=I_\Lambda(y)$ **的情形）．　此外，　在直观上还可以看出，** $h(h)$ **相对于**h **来说，等于是”常数”，所以我们还应该有**

(E.5) $E(h(h)x\,|\,h)) = h(h)E(x\,|\,h)\,.$ (1.20)

另外，还容易由定义验证

(E.6) 若随机变量x **与随机变量**h **独立，　则条件期望** $E(x\,|\,h)=Ex$.

对于多维随机变量的情形，　我们仍然有：随机变量x **关于随机向量**$\vec{h}$ **的条件期望**

$E(x\,|\,\vec{h})$ **是一个随机变量，它是**$\vec{h}$ **的函数**(**即**(E.1))**，而且满足**(E.2):

$$\forall \vec{h}\in L \Rightarrow E(I_L(\vec{h})E(x\,|\,\vec{h})) = E(I_L(\vec{h})x)\,. \qquad (1.\ 17)'$$

类似的结论还有：由(E.1)**与**(E.2) **所确定的条件期望**$E(x\,|\,\vec{h})$ **还满足**:

(E.3) $E[E(x\,|\vec{h})] = Ex$. (1. 18)’

(E.4) $E(g(\vec{h})E(x\,|\,\vec{h})) = E(g(\vec{h})x)\,.$ (1. 19)’

(E.5) $E(g(\vec{h})E(x\,|\,\vec{h})) = g(\vec{h})E(x\,|\,\vec{h})\,.$ (1. 20)’

(E.6) 若随机向量$\vec{x}$ **与随机向量**$\vec{h}$ **独立，　则** $E(f(\vec{x})\,|\,\vec{h}) = Ef(\vec{x})$ **.**

直观地我们还可以接受如下的关系：

(E.7) $E(f(\vec{x},\vec{h})\,|\,\vec{h}) = [E(f(\vec{x},\vec{y})\,|\,\vec{h})]_{\vec{y}=\vec{h}}\,.$ (1.21)

（当 $f(x,y)=g(x)h(y)$ **时，（**E.7**）显然成立．这个性质的一般情形的证明，需要用理论概率论中的典型逼近法，我们略去证明）．**

（作为特殊情形，当随机向量$\vec{x}$ **与随机向量**$\vec{h}$ **独立时，有** $E(f(\vec{\xi},\vec{\eta})\,|\,\vec{\eta}) = [Ef(\vec{\xi},\vec{y})]_{\vec{y}=\vec{\eta}}$ **.）**

(E.8) 线性 $E((ax+bV)|\vec{h})=aE(x|\vec{h})+bE(V|\vec{h})$. (1. 22)

(E.9) 平滑性: $E(E(x|h)|h,V)=E(x|h)$

$$E(E(x|h,V)|h)=E(x|h) \quad (1.23)$$

最后, 由于关于$\vec{h}$的条件期望就是在$\vec{h}$确定时的预期, 它应该具有某些最佳近似性质, 即:

(E.10) 若 $Ef(\dot{x})^2<\infty$ **,则**

$$E[f(\vec{x})-E(f(\vec{x})|\vec{h})]^2=\min_{\text{一切"好的"实函数}g}E[f(\vec{x})-g(\vec{h})]^2. \quad (1.24)$$

([**注**] **证明**:

$$E[f(\vec{x})-g(\vec{h})]^2$$
$$=E[f(\dot{x})-E(f(\dot{x})|\dot{h})]^2+E[E(f(\dot{x})|\dot{h})-g(\dot{h})]^2+2E([f(\dot{x})-E(f(\dot{x})|\dot{h})][E(f(\dot{x})|\dot{h})-g(\dot{h})])$$
$$=E[f(\vec{x})-E(f(\vec{x})|\vec{h})]^2+E[E(f(\vec{x})|\vec{h})-g(\vec{h})]^2$$
$$\geq E[f(\dot{x})-E(f(\dot{x})|\dot{h})]^2$$

在以上的推导中我们用到了: 对于任意实函数 g **有**

$$E([f(\dot{x})-E(f(\dot{x})|\dot{h})][E(f(\dot{x})|\dot{h})-g(\dot{h})])$$
$$=E[f(\vec{x})(E(f(\vec{x})|\vec{h})-g(\vec{h}))]-E[E(f(\vec{x})|\vec{h})(E(f(\vec{x})|\vec{h})-g(\vec{h}))]$$

$$=E[f(\vec{x})(E(f(\vec{x})|\vec{h})-g(\vec{h}))]-E[E([E(f(\vec{x})|\vec{h})-g(\vec{h})]f(\vec{x})|\vec{h})]$$

$$=E[f(\vec{x})(E(f(\vec{x})|\vec{h})]-E[f(\vec{x})g(\vec{h}))]-E([E(f(\vec{x})|\vec{h})f(\vec{x})]|\vec{h})-E[E([f(\vec{x}))g(\vec{h})]|\vec{h})]=0)$$

定义1．19(定义1．17的推广) **若**$Ex^2<\infty,\overrightarrow{h^{(n)}}=(h_1,\cdots,h_n)$, **则定义**

$$E(x|h_k:k=1,2,\cdots)=(L^2)\lim_{n\to\infty}E(x|\overrightarrow{h^{(n)}}). \quad (1.25)$$

2.4 **期望与方差的 Wald 等式**

本段将给出求和项数为随机变量的独立随机变量和的期望与方差。 我们先给出求和的项数随机变量与求和的项相互独立的情形的 Wald **等式**

定理1.20(独立情形的 Wald 等式) **若**$\{x_1,\cdots,x_n,\cdots\}$**为独立同分布, 且它们与另一个正整值随机变量**h**独立, 则**

$$E(x_1+\cdots+x_h)=Ex_1Eh\ , \quad (1.26)$$

$$Var(x_1+\cdots+x_h)=EhVarx_1+Var(h)Ex_1^{\ 2}. \quad (1.27)$$

证明: 利用$\{x_1,\ldots x_n\cdots\}$**与**h**的独立性, 由全期望公式我们有**

$$E(x_1+\cdots+x_h)=E(E[(x_1+\cdots+x_h)|h])=E[hEx_1]=EhEx_1,$$

$$E(x_1+\cdots+x_h)^2=E(E[(x_1+\cdots+x_h)^2\mid h])=E[hEx_1{}^2+h(h-1)(Ex_1)^2]$$

$$=EhEx_1{}^2+(Eh^2-Eh)(Ex_1)^2 .$$

从而有

$$Var(x_1+\cdots+x_h)=EhVarx_1+Var(h)Ex_1{}^2 .$$

下面再给出一般情形下的 Wald **等式**

定理１.２１（一般情形的 Wald 等式）　**若**$\{x_1,\cdots,x_n,\cdots\}$**为独立同分布，而另一个正整值随机变量**$h$**虽然并不与它们独立，但是满足如下的 Wald 条件：**

对于任意n**，事件**$\{h=n\}$**与随机变量序列**$\{x_{n+1},x_{n+2},\cdots\}$**独立**

(**其含义为：为了确定**$\{h=n\}$**这个事件发生与否，并不需要时刻**n**以后的观测**$\{x_{n+1},x_{n+2},\cdots\}$**，这种性质又称为停时性质，或称**$h$**为**$(x_n)$**停时)，那么**(1. 26), (1. 27) **仍然成立**.

证明：由 Wald **条件立即推出**$\{h<n\}$**与**$\{x_n,x_{n+1},\cdots\}$**独立，故**$\{h<n\}$**的对立事件**$\{h\geq n\}$**也与**$\{x_n,x_{n+1},\cdots\}$**独立．于是**

$$E(\sum_{n=1}^{h}x_n)=E(\sum_{n=1}^{\infty}x_nI_{\{n\leq h\}})=\sum_{n=1}^{\infty}E(x_nI_{\{h\geq n\}})=\sum_{n=1}^{\infty}Ex_nEI_{\{h\geq n\}}$$

$$=Ex_1\sum_{n=1}^{\infty}P(h\geq n)=Ex_1Eh .$$

$$E(\sum_{n=1}^{h}x_n)^2=E(\sum_{n=1}^{\infty}x_n{}^2I_{\{n\leq h\}})+E(\sum_{\substack{n,m=1\\n\neq m}}^{\infty}x_nx_mI_{\{n.m\leq h\}})$$

$$=Ex_1{}^2Eh+(Ex_1)^2\sum_{\substack{n,m=1\\n\neq m}}^{\infty}P(h\geq n,m)$$

再注意到

$$\sum_{\substack{n,m=1\\n\neq m}}^{\infty}P(h\geq n,m)=\sum_{n,m=1}^{\infty}P(h\geq n,m)-Eh=E\sum_{n,m=1}^{\infty}I_{\{n\leq h\}}I_{\{m\leq h\}}-Eh$$

$$=Eh^2-Eh .$$

由此可知 Wald **等式成立**.

3　统计简要

3. 1　用样本作矩估计

定义１.２２ 独立同分布的随机变量$x_1,\cdots,x_n$ （或随机向量）称为大小为 n 的一组随机

样本. 样本$x_1,\cdots,x_n$的平均 $\overline{x}\overset{\Delta}{=}\dfrac{x_1+\cdots+x_n}{n}$ 可用来估计Ex_1．样本$x_1,\cdots,x_n$的修正方差

$s^2\overset{\Delta}{=}\dfrac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n-1}$ 可用来估计$Varx_1$. 一般地, $\overline{x^k}\overset{\Delta}{=}\dfrac{x^k{}_1+\cdots+x_n{}^k}{n}$ 称为$Ex_1{}^k$的矩估计.

矩估计并不依赖于分布, 故有很好的稳健性与普适性. 这是它的长处. 当然, 如果知道了x_1的分布的信息, 则可以更多地利用它们去构造更有效的估计方法, 例如, 可以用最大似然估计法, 最小二乘法，或 Bayes 估计法.

矩估计都是无偏的, 即: $E\overline{x}=Ex_1$, $Es^2=Varx_1$, 而且也是相合的, 即: 当

$n\to\infty$ 时有 $\overline{x}\xrightarrow{p}Ex_1$, $s^2\xrightarrow{p}Varx_1$,

用相合估计作为参数的估值的含义在于: 以可以忍受的小概率失败的风险来换取能以大概率得到成功的效果.

3. 2 最大似然估计

定义１.２３ 令$p(x,q)$为$x_1,\cdots,x_n$们的概率函数(离散型情形)，或概率分布密度(连续型情形), 其中q是未知参数, 它与x也都可以是向量. 记

$$L(q)=\prod_{i=1}^{n} p(x_i,q),$$

称为样本$x_1,\cdots,x_n$的似然函数.

若x为连续型随机变量，而h为离散型随机变量(为了使记号更简便，我们不妨假定它的取值为 $j=1,2,\cdots$)，设随机向量(x,h)的联合分布依赖于参数q．于是此联合分布可表示为：

$$P(x\le x,h=j)=\int_{-\infty}^{x}p_j(u,q)du\ ,$$

此时x具有分布密度 $p(x,q)=\sum_j p_j(x,q)$， 而$P(h=j)=\int_{-\infty}^{\infty}p_j(u,q)du$．

对于随机向量(x,h)的n个随机(独立同分布的)样本$(x_i,h_i),(i\le n)$，其似然函数为

$$L(q)=f(\vec{x},\vec{h}),$$

其中

$$f(\vec{x},\vec{y})=\prod_{i=1}^{n}\frac{d}{dx_i}P(\xi\le x_i,\eta=y_i)=\prod_{i=1}^{n}p_{y_i}(x_i)\text{，}$$

而 y_i 是 $1,2,\cdots$ 中的一个数．

定义１．２４ 如果样本 $x_1,\cdots,x_n$ 的一个函数 $\hat{q}_M=\hat{q}_M(x_1,\cdots,x_n)$ 满足:

$L(\hat{q}_M)=\sup_q L(q)$， 则称 $\hat{q}_M$ 为 q 的最大似然估计.

最大似然估计是似然函数整体最大值的位置，通常用局部极大值的位置作为其粗略的近似．要减少估计的随机误差（方差），就需要进行整体优化，为此要借助于数值方法或随机模拟.

3. 3 线性模型的最小二乘估计及其推广

设

$$\dot{x}\overset{\Delta}{=}\begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix}=\begin{pmatrix}a_{11}&\cdots&a_{1k}\\ \vdots&\vdots&\vdots\\ a_{n1}&\cdots&a_{nk}\end{pmatrix}\begin{pmatrix}q_1\\ \vdots\\ q_k\end{pmatrix}+\begin{pmatrix}e_1\\ \vdots\\ e_k\end{pmatrix}\overset{\text{记为}}{=}A\dot{q}+\dot{e}\quad ,$$

其中 $n\ge k$， A 为已知的满秩矩阵（即秩为 k ），$\vec{q}$ 为未知的待估参数，需要通过随机向量 $\vec{x}$ 的观测值来给出 $\vec{q}$ 的估计 $\hat{q}$.

(1) 简单情形　　设 $e_1,\cdots,e_k$ 为独立同分布，它们的数学期望为零，方差为 s^2.

取 $\vec{q}$ 的估计为 $\hat{q}=(A^TA)^{-1}A^T\dot{x}$．那么，它满足

$$E\,|\dot{x}-A\hat{q}\,|^2=\inf_{\dot{q}}E\,|\dot{x}-A\dot{q}\,|^2,$$

其中 $|\vec{x}|^2=\sum_{i=1}^{n}x_i^{\,2},\vec{x}=(x_1,\cdots,x_n)$.

定义１．２５ $\hat{q}$ 称为矛盾方程 $A\vec{q}=\vec{x}$ 的最小二乘解， 在应用中则称为参数 $\vec{\vartheta}$ 的最小二乘估计． 显见它是 $\vec{\vartheta}$ 的无偏估计.

(2) 推广情形　　设随机向量 $(e_1,\cdots,e_k)$ 服从数学期望为零的正态分布， 其方差矩阵为正定矩阵 Σ .

定义１．２６ 取 $\vec{q}$ 的估计为 $\hat{q}=(A^T\Sigma^{-1}A)^{-1}A^T\Sigma^{-1}\dot{x}$， 它称为 $\vec{\vartheta}$ 的广义最小二乘估计．显见它也是 $\vec{\vartheta}$ 的无偏估计.

(3) **加权最小二乘估计** **例如，可以用**$x^{(a)}=\begin{pmatrix} a^{n-1}x_1 \\ \vdots \\ a x_{n-1} \\ x_n \end{pmatrix}(0<a<1)$ **代替**$\vec{x}$**．这时，矛盾方程**$A\vec{q}=\overrightarrow{x^{(a)}}$**的最小二乘解**$\hat{q}^{(a)}$**，就称为矛盾方程** $A\vec{q}=\vec{x}$ **的**a**加权最小二乘解．使用它的考虑是**: **对历史较早的资料**, **相对地重要性要小一些，所以在利用时取较小的权重**.

习题 1 概率论复习题

1\. (1) **判断下式是否成立（填＋或－）：**$P(AB\mid A)=P(B\mid A)$ **（ ）．**

(2) (**填空**) **若** A,B **互不相容，则** $P(C\mid A\cup B)=(\quad)P(C\mid A)+(\quad)P(C\mid B)$**．**

(3) (**填空**) $P(B\mid(A-B))=$_____**，** $P(A\mid(A-B))=$_____**．**

(4) **设随机事件** A,B **的示性函数为** I_A,I_B. **求** (I_A,I_B) **的联合分布，**$D(I_A),\mathrm{cov}(I_A,I_B)$ **及** I_A **与的** I_B **相关系数．**

(5) (**填空**) X **为随机变量，**A **为随机事件，**B_k **们两两不相交**, **且**

$$\bigcup_k B_k=\text{必然事件}\,\Omega\text{，则}\ \sum_k P(B_k\mid A)E(X\mid AB_k)=(\quad).$$

2．证明：（1） $|P(AB)-P(A)P(B)|\le\frac{1}{4}$**；（2）** $|P(AB)-P(BC)|\le 1-P(AC)$**；**

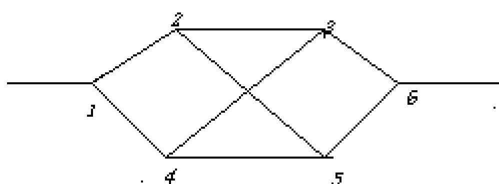
（3） $P(A_1\cdots A_n)\ge 1-\sum_{k=1}^{n}(1-P(A_k))$**；**

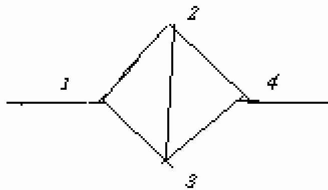
（4）若 $P(A)=p,P(B)=1-p$**，则** $P(A^cB^c)=P(AB)$**；**

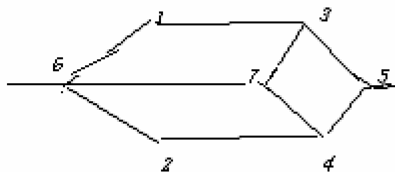
（5）若 $P(B)P(B^c)>0$**，**$P(A\mid B)=P(A\mid B^c)$**，则** A,B **独立**

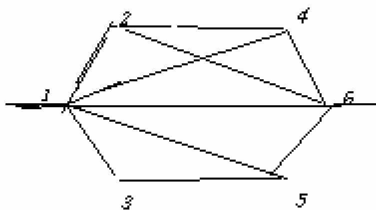
3．（1） 由 $P(A\mid B)>P(A)$ **推出** $P(B\mid A)>P(B)$**，解释其含义；**

（2） 设第 i **个元件的可靠性为** p_i**， 且元件间相互独立，分别求以下系统的可靠度：**









（3） 化简 $P(AB \mid A\cup B)$，$P(AB \mid A\cup B\cup C)$，$Cov(aX+b, cX^{2}+d)$

4．设 X 为取非负整数值的随机变量，证明：

（1） $EX=\sum_{n=1}^{\infty}P(X\geq n)$

（2） $DX=2\sum_{n=1}^{\infty}nP(X\geq n)-EX(EX+1)$

5．设随机变量 X 与 Y 独立，且方差有限，则

（1） $D(XY)=DXDY+(EX)^{2}DY+(EY)^{2}DX$

（2） $D(XY)\geq DXDY$ 。

6．设$x_1,x_2\cdots,x_{n+m}$ $(n>m)$相互独立同分布且具有有限方差，试求 $X=\sum_{k=1}^{n}x_k$ 与$Y=\sum_{k=1}^{n}x_{m+k}$ 的相关系数。

7. $X_1,\cdots,X_n$独立同分布， 其分布为$\begin{pmatrix}0 & 1\\ q & p\end{pmatrix}$. $S_n=X_1+\cdots+X_n$． 问

$$P(X_1=1\,|\,S_n=j)\ P(X_2=0\,|\,S_n=j)\text{与}P(X_1=1,X_2=0\,|\,S_n=j)$$

那个大？其直观的含义是什么？

8．$X_1,\cdots,X_n$独立同分布， 其分布为$\begin{pmatrix}-1 & 1\\ q & p\end{pmatrix}$. $S_n=X_1+\cdots+X_n,S_0=0$，试求：$E(S_3\,|\,S_2)$ 的分布列，并证明$E(S_{n+1}\,|\,S_n)=S_n+(p-q)$。

9．设随机变量 *X* 服从参数为 l 的 Poisson 分布，求$E\,|\,X-l\,|$。

10．设 *X* 与 *Y* 分别服从参数为 l_1 和 l_2 的 Poisson 分布，且 *X* 与 *Y* 独立．证明在给定 *X+Y* 下，*X* 的条件分布是二项分布。

11．设独立试验的次数 *X* 服从参数为 l 的 Poisson 分布， 其中每次试验成功的概率为 *p*，不成功的概率为$1-p$．求试验成功次数的概率分布。

12．设随机变量 *X* 服从参数为 l 的 Poisson 分布，且 l 也是一随机变量，服从两点分布：

$P(l=l_1)=p,P(l=l_2)=1-p$，求 X 的分布列。

13． 设随机变量 X 具对称的密度函数，即$f(-x)=f(x)$，证明对任意的$a>0$，有

(1) $F(-a)=1-F(a)=\frac{1}{2}-\int_0^a f(x)dx$；

(2) $P(|X|<a)=2F(a)-1$；

(3)．$P(|X|>a)=2(1-F(a))$

14．设随机变量 k 服从（0，5）上的均匀分布。求方程

$$4x^2+4kx+k+2=0$$

有实根的概率。

15．设随机变量 X 服从正态分布$N(m,s^2)$． 求$E\,|\,X-m\,|$．

16. 零件强度 *X*～$N(48,2^2)$．若 $X\geq 46.32$ 为合格品，否则为次品．检验方案为：第一次任取 3 个，若 3 个都合格，则接收该批零件；若至少有 2 个次品，则拒收该批；否则再抽第二次，也是 3 个。重复上述

步骤，至作出接收或拒绝决定为止。

(1) 求由第一次取样而接收该批零件的概率；

(2) 求该批产品被接收的概率；

(3) 记 N 为在作出决定时的抽取次数, 求 N 的分布.

17. 如果随机变量x 有分布密度 $p(x) = b \cdot e^{ab} \dfrac{1}{\sqrt{2p \cdot x^3}} e^{-\frac{1}{2}(\frac{b^2}{x}+a^2 x)} I_{[0,\infty)}(x)$， 则称为逆 Gauss 分布， 记为 $x \sim IG(b,a)$。 假定x_1, x_2独立，$x_i \sim IG(b_i,a)$。证明$x_1 + x_2 \sim IG(b_1+b_2,a)$.

18. 设 $X \sim N(m, s^2)$, 求实数a使$E(X-a)^4 = \min_b E(X-b)^4$

19．设随机变量 X 的概率密度函数为 $f(x)$，令$h(a) = E|X-a|$，试证明：当a满足

$P(X \le a) = \dfrac{1}{2}$时（此时称 a 为 X 的中位数），$h(a)$ 达到最小。

20．设$t \sim Exp_l$，求 $t \wedge t$ 的分布函数.

21．设$x \sim Exp_l, h \sim Exp_l, x, h$ 独立， 求$P(x > h)$并求$x \wedge h$ 及$(x-h)^+$的分布.

再记 $X = x \wedge h, Y = |x-h|, Z = I_{\{X \le Y\}}$. 问 X, Y 是否独立 ？ Y, Z 是否独立 ？

22．设x, h，$V \sim N(0,1), x, h$，V 独立，求 $E(x \wedge h), E(x \vee h)$并求$x \wedge h$ 及 $x \vee h$ 的分布及

$P(x+h > 2V)$，$P(\dfrac{x+h}{2} - \dfrac{x+h+V}{3} > 0)$

23．设 $F(x)$ 为分布函数，$\overline{F}(x) = 1 - F(x)$．证明 $F(x)$ 是指数分布的充要条件为

$$\overline{F}(t+s) = \overline{F}(s)\overline{F}(t)$$

24．设随机向量(X,Y)服从椭圆$\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} \le 1 \quad (a > 0, b > 0)$上的均匀分布．求其相关系数． 问 X 与 Y什么时候独立 ？

25．设随机变量 X的概率密度为

$$f(x) = \frac{1}{2} e^{-|x|}, \quad (-\infty < x < \infty)$$ （称为 Laplace 分布）.

(1)．求 X的数学期望EX 和方差DX ；

(2)．求 X与$|X|$ 的协方差． 又 X与$|X|$ 是否不相关？ 是否相互独立？ 为什么？

26．设随机变量 X服从$N(0,1)$ 分布．求 X 与 X^n （n为正整数）的协方差与相关系数。

78. 设$x \sim \exp_l$，给定$c > 0$，求在$x > c$)条件下， x 的条件分布密度。

28．设随机变量$h \sim U[0,a]$， $X \sim U[h,a]$．证明随机变量 $E(X|h) \sim U[\dfrac{a}{2}, a]$。

29．假设 $E(X\mid Y)=EX$ ，证明 X 与 Y 不相关．举例说明其逆命题不成立。

30．求 Pareto 分布（密度 $f(x)=\dfrac{a\,l^a}{(l+x)^{a+1}}I_{(-l,0]}(x)$)的期望，方差，矩，失效率函数

31．若随机变量 *X* 与 *Y* 相互独立，且分别服从参数为 l 与 m 的指数分布．定义随机变量

$$Z=\begin{cases}1 & \text{当}X\le Y\\ 0 & \text{当}X>Y\end{cases}$$

（1）求条件概率密度 $f_{X|Y}(x\mid y)$ ；

（2）求 Z 的分布列和数学期望．

32．设 $X_1,\cdots,X_n$ 为 n 个具有密度 $f(x)$ (分布函数记为 $F(x)$)的独立同分布随机变量．令 $X_{(1)}$ 为它

们中最小的值，即 $X_{(1)}=\min_{k\le n}X_k$ ， $X_{(2)}$ 为它们中第二个小的值，…，$X_{(n)}$ 为它们中最大的值．

（1）． 证明 $X_{(k)}$ 具有密度 $kC_n^k f(x)F(x)^{k-1}(1-F(x))^{n-k}$ ．

（2）． 若独立同分布的 $X_1,\cdots,X_n\sim U[0,1]$ ，求 $EX_{(k)},Var(X_{(k)})$ 及相关系数 $r(X_{(1)},X_{(n)})$ ．又问 $Var(X_{(1)}),\cdots,Var(X_{(n)})$ 中那个最小， 那个最大 ？

（3） 证明当 $n\to\infty$ 时，$X_{(n)}$ 按概率收敛到 1，且 $n(1-X_{(n)})$ 的分布收敛到 Exp_1 ．

33. (1). 证明 $Var(Y)\ge Var(Y\mid X)$, 其中 $Var(Y\mid X)=E[(Y-E(Y\mid X))^2\mid Y]$ 称为 Y 关于 X 的条件方差.

(2) 在 (X,Y) 服从二维正态分布时, 求证 $Var(Y\mid X)=r^2Var(Y)$, 其中 r 是 X 与 Y 的相关系数.

(3). 作为(1)的推广, 证明 $Var(Y\mid X)\ge Var(Y\mid X,Z)$.

(4) $Cov(Y,Z)=E(Cov[(Y,Z)\mid X])+Cov(E(Y\mid X),E(Z\mid X))$ ，

其中 $Cov((Y,Z)\mid X)=E[(YZ)^2\mid X]-E(Y\mid X)E(Z\mid X)$ 是随机变量 Y 和随机变量 Z 关于随机变量 X 的条件协方差．作为特殊情形，我们有 $Var(Y)=E[Var(Y\mid X)]+Var(E(Y\mid X)$ ．

34. 独立地重复扔一个正面出现的概率为 p 的有偏钱币 n 次. 令 Y 为其中出现正面的次数, $X_k(k<n)$ 为第 k 次出现正面时的抛扔次数. 求证

$$P(X_1 = i_1, \cdots, X_m = i_m \mid Y = m) = \frac{1}{C_n^m}, (m < n, i_1 < \cdots < i_m).$$

35. X, Y **相互独立，** $X \sim N(m, s^2), Y \sim Poisson_l$，**求** $\frac{X}{Y+1}$ **的分布．**

36．设$[0,\infty)$ **上分布函数** $F(t)$ **具有密度** $f(t)$．**定义** $l(t) = -\frac{f(t)}{1-F(t)}$，**称为失效率函数．证明**

$$P(t < x < t+h \mid x > t) = l(t)h + o(h)，\ F(t) = 1 - e^{-\int_0^t l(s)ds}.$$

再分别求指数分布，Weibull 分布和 Logistic 分布的失效函数．

37．求 $P(X \le x \mid a < X \le b)$ **和它的失效率（截尾失效率）．落实到** $X \sim N(m, s^2)$ **情形。**

38. **设一批零件的长度** X～N$(50, 1^2)$．**当** $|X - 50| \le 2$ **为合格品，否则为次品。求将产品中不合格元件全部剔除后，剩下的零件的长度的分布．**

39．若系统由两种独立工作的设备串联而成．它们的寿命分别为参数 l, m **的指数分布，又它们各有备件** $n-1$ **件与** $m-1$ **件．求在这些备件的支持下，系统的平均寿命．**

40．设$h \sim U[0,1]$],　　**记**

$$h_n = \sum_{k=0}^{2^n-1} \frac{k}{2^n} I_{\{\frac{k}{2^n} \le h < \frac{k+1}{2^n}\}}.$$

(1)　**求** h_n **的分布；**

(2)　**给定** $h_1 = 1/2$ **下，** h_2 **的条件分布律**.

41. **设** A, B **为随机事件．　已知** $P(A), P(B), P(AB)$ **，求** $E(I_B \mid I_A)$，$E(I_{A \bigcup B} \mid I_A)$ **及** $E(I_{A \bigcap B} \mid I_A)$，**并分别求它们的分布．**

42．　若$y(t) \overset{\Delta}{=} Ee^{tX} < \infty$，　**证明以下的 Chernoff 不等式：**$\forall t$

$$P(X > a) \le e^{-(at - y(t))}.$$

再由此推出

$$P(X > a) \le e^{-\sup_t (at - y(t))}.$$

43．设 $X_1, X_2, \cdots, X_n$ **独立同分布，其密度函数为**

$$f(x) = \begin{cases} 2x & 0 < x < 1 \\ 0 & \text{其它} \end{cases}$$

记 $Z_n = n[1-\max(X_1,\cdots,X_n)]$ 的分布函数为 $G_n(x)$．证明：

$$\lim_{n\to\infty} G_n(x) = 1-e^{-2x}.$$

44. 设 $\{X_n\}$ 为独立同分布的随机变量序列，服从 Poisson $_l$．证明可用积分 $\frac{1}{\sqrt{2p}}\int_{-\infty}^{a} e^{-\frac{t^2}{2}}dt$ 来近似计算

$$\sum_{0\le k\le u}\frac{(nl)^k e^{-nl}}{k!} \quad (\text{其中 } a=\frac{u-nl}{\sqrt{nl}}) \quad .$$

45. $\cos^2 q \cdot e^{-2q^2-q\cdot i} e^{(e^{iq}-1)}$ 是否是随机变量的特征函数？ 若回答"是"，则问这个随机变量是什么．

46. 下列等式可以用来简化某些矩（例如从低中心矩求高一阶的原点矩）的计算

（1） 设 $X \sim N(m,s^2)$， $E|g'(X)|<\infty$． 用分部积分证明 Stein 引理：

$$E[g(X)(X-m)] = s^2 E(g'(X).$$

再用 $g(x)=x^2$ 证明 $EX^3 = m^3+3m\cdot s^2$．

（2） 设 $X \sim Poisson_l$，$Eg(X)$ 有限． 证明 Hwuang 引理：

$$E[l\cdot g(X)] = E[Xg(X-1)].$$

再用 $g(x)=x^2$ 证明 $EX^3 = l^3+3l^2+l$ ．

（3） 设 $X \sim$ 参数为 (r,p) 的负二项分布． 证明

$$E[(1-p)g(X)] = E\left[\frac{X}{r+X-1}g(X-1)\right].$$

（4） 设 $X \sim G(a,l)$． 证明 $E[g(X)(X-\frac{1}{l}] = \frac{1}{l}E[Xg'(X)]$．

（5） 设 $X \sim Beta(a,b)$． 证明 $E[g(X)(b-(a\cdot -1)\frac{1-X}{X}] = E[(1-X)g'(X)]$．

47. 设随机变量服从指数型分布，其密度为 $p(x)=C(\vartheta)h(x)e^{\sum_{k=1}^{m}\varphi_k(\vec{\vartheta})\cdot T_k(x)}$． 证明

$$E(\sum_{i=1}^{m}\frac{\partial j_i(\vec{J})}{\partial J_j}T_i(x)) = -\frac{\partial \log C(\vec{J})}{\partial J_j}.$$

（这个等式的意义在于把左边的积分计算简化为右边的微分运算）．

48. 设 $EX=m$ ，$g(x)$ 非降．证明 $E(g(X)(X-m)\ge 0$．并解释其含义．

49. 若 $g(x),h(x)$ 都是非增函数，或都是非降函数．证明 $E[g(X)h(X)]\ge Eg(X)Eh(X)$， 并解

释其含义．又若它们一个非增，另一个非降，则结果如何 ？

50. 设随机变量 X 服从 $[0,1]$ 均匀分布，而在随机变量 $X=x$ 条件下，随机变量 Y 的条件分布为 $N(x,x^2)$． （1）求 $EY,Var(Y),Cov(X,Y)$． （2）证明随机变量 $\frac{Y}{X}$ 与 X 独立．

51. 设随机变量 X 服从 $[0,1]$ 均匀分布，而在随机变量 $X=x$ 条件下，随机变量 Y 的条件分布为二项分布 $B(n,x)$．（1）求 $EY,Var(Y)$． （2）求随机变量 Y 的分布．

52. 设随机变量 X 服从 $G(a,l)$ 分布，而在随机变量 $X=x$ 条件下，随机变量 Y 的条件分布为 $Poisson_x$． （1）求 $EY,Var(Y)$． （2）求随机变量 Y 的分布．

53. 设随机变量 X 服从 $G(a,l)$ 分布，而在随机变量 $X=x$ 条件下，随机变量 N 的条件分布为 $Poisson_x$，又在随机变量 $N=n$ 的条件下．随机变量 Y 的条件分布为二项分布 $B(n,p)$。

（1）证明 $Var(Y)=EX+Var(X)$． （2）求随机变量 Y 的分布．

54. 设随机变量 X 服从 $Beta(a,b)$，而在随机变量 $X=x$ 条件下，随机变量 Y 的条件分布为二项分布 $B(n,x)$．证明 $Var(Y)=nEX(1-EX)+n(n-1)Var(X)$．（$Y$ 的分布称为 $Beta-$二项分布）．

55. 设随机变量 X 服从 $Beta(a,b)$ 分布，而在随机变量 $X=x$ 条件下，随机变量 Y 的条件分布为 $Pascal(r,x)$．求 $EY,Var(Y)$．（Y 的分布称为 $Beta-Pascal$ 分布）．

56. 设随机向量 $(x_1,\cdots,x_m)$ 服从参数为 $(p_1,\cdots,p_m)$ 的多项分布．求协方差 $Cov(x_i,x_j)$，相关系数 r_{x_i,x_j} 以及在 $X_i=k_i$ 条件下，X_j 的条件分布．

57. 假定 f_1,f_2,g_1,g_2 为 4 个分布密度，它们的数学期望分别为 m_1,m_2,n_1,n_2．今有随机向量 (X,Y)，其联合分布密度为 $l\cdot f_1(x)g_1(y)+(1-l)f_2(x)g_2(y)\quad(0<l<1)$．求 $Cov(X,Y)$，又问 X,Y 何时相互独立 ？

58. 求证 n 个独立 $U[0,1]$ 随机变量的乘积的分布密度是 $\frac{(\ln x)^n}{n!}I_{[0,1]}(x)$

59 设 $x\sim\begin{pmatrix}0 & 1\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}$，$h\sim U[0,1]$，且相互独立。分别求 $x+h$，xh 的分布

60。设 X,Y 独立同分布。求证 $P(a<X\wedge Y)\le b)=P(X>a)^2-P(X>b)^2$

61.设随机变量 $t_1,\cdots,t_n$ **相互独立，且** $t_k \sim Exp_{l_k} \quad (k \le n)$ **.证明**

$$P(t_j = \max_{k\le n} t_k) = \frac{l_k}{l_1+\cdots+l_n}.$$

62．设随机变量 $X \sim U[-p,p], S_n = \sum_{k=1}^{n} Cos(kX)$。 **证明** $\frac{S_n}{n} \xrightarrow{p} 0(n\to\infty)$.

（用 Chebyshev 不等式）.

63．设 X,Y,Z,W **是独立同分布的标准正态随机变量。 证明**

$$P(a\sqrt{X^2+Y^2} > b\sqrt{Z^2+W^2} = \frac{b^2}{a^2+b^2}$$

64．分别对于随机变量 $X \sim N(m,s^2), X \sim Poisson_l, X \sim U[-1,1]$， **求**

$E(CospX)), Var(Cos(pX))$ **（利用特征函数）.**

65．将出现在空间区域 Λ **中的随机的粒子的数目记为随机变量** X_Λ。 **假设** $X_\Lambda \sim Poisson_{a|\Lambda|}$， **其中**

a **是正常数，** $|\Lambda|$ **表示空间区域** Λ **的体积．取定其中一个粒子后，与它最近的粒子对于它的距离记**

为 h **．求证随机变量** h **服从 Weibull 分布， 即它的分布函数为** $F_h(x) = 1-e^{-a\frac{4}{3}px^3}$ **（注： 如果**

随机粒子出现的空间不是三维空间， 而是 d **维空间， 则相应地有** $F_h(x) = 1-e^{-a\frac{p^{\frac{d}{2}}}{\Gamma(\frac{d}{2}+1)}x^d}$ **）.**

66．随机变量 X,Y **独立，具有相同的分布** $N(m,s^2)$。 **求** $E(X\wedge Y)$

67．设随机变量 X,Y **满足：** $EX = EY = 0, Var(X) = Var(Y) = 1, Cov(X,Y) = r$ 。

(1) 求 $E(X\pm Y)^2$ **；　(2) 证明如下的 Chebyshev 不等式：**

$$E(X^2 \vee Y^2) = \frac{1}{2}E(X^2+Y^2+|X^2-Y^2|) \le 1+\sqrt{1-r^2}\text{。}$$

68．利用 67 题的结论，证明二维 Chebyshev 不等式：

$$P(|X-EX| \ge eVar(X), |Y-EY| \ge eVar(Y)) \le \frac{1+\sqrt{1-r^2}}{e^2}\text{。}$$

你能否把这想法推广到三维情形 ？

69．. **设** $F(x), G(y)$ **是两个分布函数．记**

$$F_1(x,y) = ((F(x)+G(y)-1)^+, F_2(x,y) = \min(F(x),G(y)),$$

其中记号 $z^+ = zI_{[0,\infty)}(z)$ **. 证明** $F_1(x,y), F_2(x,y)$ **都是二元分布函数，且满足**

（1） $F_1(x,\infty) = F_2(x,\infty) = F(x)$**，** $F_1(\infty,y) = F_2(\infty,y) = G(y)$ **；**

（2） **若另有二元分布函数** $F(x,y)$ **，也满足** $F(x,\infty) = F(x)$**，** $F(\infty,y) = G(y)$ **，则**

$F_1(x,y) \le F(x,y) \le F_2(x,y)$.

龚光鲁，钱敏平著 应用随机过程教程及其在算法与智能计算中的应用
清华大学出版社, 2003

第２章 随机样本生成法

1 一维随机数

随机变量（或随机向量）的样本简称为随机数. 由于在统计中常用的是独立样本列，所以我们不妨假设随机数之间都是独立的．生成随机数的方法，也称为随机数的取样法(Sampling)．

1. 1 均匀随机变量的计算机模拟

定义２.１ 在$[0,1]$上均匀分布的随机变量的独立样本称为均匀随机数($U[0,1]$ 随机数).

在计算机上产生的称之为”伪随机数”的数列, 是一种具有非常长周期的, 且能通过数理统计中的独立性与均匀性假设检验的数列. 实践证明伪随机数是均匀随机数的一种可行的近似. 这种伪随机数虽然并不是独立同分布的$U[0,1]$ 随机变量的样本, 而是在$[0,1]$中取值的周期数列, 但是由于它可以像均匀随机数一样地通过数理统计中的独立性与均匀性假设检验, 而且它的周期非常长, 以至在计算机实际运算过程中不会出现重复, 所以在实际计算中它能很好地替代均匀随机数.

最普遍用以产生伪随机数的方法是同余法. 典型的例子如下:

$$y_{n+1} = 5^{13} y_n (\mathrm{mod} \quad 2^{36}), y_0 = 1, x_n = y_n \cdot 2^{-36} \quad (\text{周期约为} 2\cdot 10^{10})\text{；}$$

$$y_{n+1} = 5^{17} y_n (\mathrm{mod} \quad 2^{42}), y_0 = 1, x_n = y_n \cdot 2^{-42} \quad (\text{周期约为} 10^{12})\text{；}$$

$$y_{n+2} = y_{n+1} + y_n (\mathrm{mod} \quad 2^{44}), y_0 = 0, y_1 = 1, x_n = y_n \cdot 2^{-44} \quad (\text{周期约为} \frac{1}{4}\cdot 10^{14})\text{．}$$

（关于伪随机数， 可参见：现代数学手册，随机数学卷，第 10 篇，孙嘉阳， 石坚，丛树铮，徐映波编 蒙特卡罗法．华中科技大学出版社，2000 年）．

1. 2 分布函数$F(x)$的随机数

命题２.２（反函数方法）分布函数为$F(x)$的独立随机变量列的样本，称为 $F(x)$ 随机数．若$F(x)$严格单调递增, x 是均匀随机数, 则$F^{-1}(x)$是$F(x)$随机数, 其中F^{-1}为F 的反函数.

证明 $P(F^{-1}(x) \le x) = P(x \le F(x)) = F(x)$. ?

命题２．３ 设随机变量x只取有限个值，其分布为 $x \sim \begin{pmatrix} x_1 & \cdots & x_n \\ p_1 & \cdots & p_n \end{pmatrix}$. 把$[0,1]$分为$n$个不交子区间, 使第$i$个区间$J_i$的长度为 p_i. 任取均匀随机数U , 则

$$\sum_{i=1}^{n} x_i I_{J_i}(U)$$

就是一个x 随机数（它的意思是：如果U 落入J_i，就取$x = x_i$）.

在统计再抽样中的应用

在样本组中再抽样，或者由样本作的参数估计代替分布中的未知参数后，所得到的分布的随机取样，统一称为 Bootstrap 方法．具体地说有如下两种方法

（1）非参数 Bootstrap 方法．设自一个未知方差的分布取样 $X_1,\cdots,X_n$ （不是计算机模拟取样，而是人工取样）如果要作方差的区间估计，就需要知道方差估计 $\hat{s}^2$ 的方差 $Var(\hat{s}^2)$．一般$Var(\hat{s}^2)$很不好求，需要对它用再抽样进行估计．为此可将样本分布 $\begin{pmatrix} X_1 & \cdots & X_n \\ \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}$作为离散随机变量的分布，独立地取样$N$ 次，每次独立地取样m 个．设从第k 次的m 个样本值得到方差的估计 $\hat{s}^2_k$ $(k \le N)$，将此N 个的平均记为 $\overline{\hat{s}^2}$ ，最后用 $\hat{\hat{s}}^2 = \frac{1}{N-1}\sum_{k=1}^{N}(\hat{s}^2_k - \overline{\hat{s}^2})^2$ 估计$Var(\hat{s}^2)$．

此法可以用于一般未知参数的方差估计．

（2）参数 Bootstrap 方法． 设自一个带有未知参数$(J_1,\cdots J_l)$的分布 $p(x,J_1,\cdots J_l)$ 的样本 $X_1,\cdots,X_n$（不是计算机模拟取样，而是人工取样）得到未知参数的估计$(\hat{J}_1,\cdots,\hat{J}_l)$后，对分布 p（$x,\hat{J}_1,\cdots,\hat{J}_l$)用计算机模拟取样．独立地取样 N 次，每次独立地取样 m 个．其它与（1）相同．

注意，计算机模拟取样只能对已知的分布施行，对于含未知参数的分布，只能作普通的人工取样．以上的两种再抽样方法，补充了人工取样采样量的限制．因为计算机模拟取样既快速又经济．

1．3 **正态随机数**

$N(0,1)$ 随机数称为标准正态随机数． 生成标准正态随机数有一个比反函数的方法更为简单的实践方法, 就是利用中心极限定理. 设$h_1,\cdots,h_{12}$为均匀随机数（它们是独立的），由中心极限定理，可以认为$x = h_1 + \cdots + h_{12} - 6 \approx N(0,1)$，即用$x = h_1 + \cdots + h_{12} - 6$ 近似地作为标准正态随机数. 在实际计算中h_i $(1 \le i \le 12)$ 们还应该用伪随机数代替.

命题２.４ （生成标准正态随机数的 Box-Muller 方法） 取两个独立的均匀随机数

h_1, h_2，令

$$x_1 = \sqrt{-2\ln h_1}\cos(2p h_2),$$

$$x_2 = \sqrt{-2\ln h_1}\sin(2p h_2).$$

则x_1, x_2为相互独立的标准正态随机数.

证明 令$h_1, h_2 \sim U[0,1]$且独立，则$1-h_1, h_2$也是独立的$U[0,1]$随机变量．于是由命题**2**.2，$r \overset{\Delta}{=} F^{-1}(1-h_1) = \sqrt{-2\ln h_1} \sim$分布函数$F(r) = 1-e^{-\frac{r^2}{2}}$，$J \overset{\Delta}{=} 2p h_2 \sim U[0,2p]$，且相互独立． 而$x_1 = r\cos J, x_2 = r\sin J$，易见它们是独立的标准正态随机数.

命题2.5 （生成标准正态随机数的 Marsaglia 方法） 设(X,Y)为单位圆上的均匀随机数．则

$$\begin{pmatrix} x \\ h \end{pmatrix} \overset{\Delta}{=} \sqrt{\frac{-2\ln(X^2+Y^2)}{X^2+Y^2}}\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right).$$

（提示 将直角坐标(X,Y)转换为极坐标(R,J)）.

一般正态随机数的生成 若x为标准正态随机数，则显见$sx+m$为$N(m,s^2)$随机数.

1.4 Poisson 随机数

下述结论给出了利用伪随机数生成 Poisson 随机数的方法。

命题2.6 设$U_1, U_2, \cdots$是相互独立的$[0,1]$均匀随机数． 若

$$\prod_{k=1}^{n+1} U_k < e^{-\lambda} \le \prod_{k=1}^{n} U_k，\quad 则定义 \quad N = n. \quad 那么 \quad N \sim Poisson_{\lambda}.$$

证明 令$T_k \overset{\Delta}{=} -\ln U_k \sim \exp_1$．在指数流与 Poisson 过程的关系（参见第 3 章）中取参数为 1，取时间t为 l 即得.

1.5 混合分布随机数

对于权重为 $p_1, \cdots, p_n$ （和为 1 的n个正数）的混合分布随机数，我们有

命题2.7 设$U \sim U[0,1], 0 = t_0 < \cdots < t_n = 1, t_i - t_{i-1} = p_i (i \le n)$，$F_i(x)(i \le n)$为分布函数，那么

$$\sum_{i=1}^{n} F_i^{-1}\left(\frac{U-t_{i-1}}{p_i}\right) I_{(t_{i-1},t_i]}(U) \sim \text{分布函数为} \sum_{i=1}^{n} p_i F_i \text{ 的混合分布.}$$

证明

$$P(\sum_{i=1}^{n} F_i^{-1}(\frac{U-t_{i-1}}{p_i})I_{(t_{i-1},t_i]}(U) \le x) = \sum_{i=1}^{n} P(F_i^{-1}(\frac{U-t_{i-1}}{p_i})I_{(t_{i-1},t_i]}(U) \le x, U \in (t_{i-1}, t_i])$$

$$= \sum_{i=1}^{n} P(t_{i-1} < U \le t_{i-1} + p_i F_i(x)) = \sum_{i=1}^{n} p_i F_i(x) .$$?

在实际计算中，应该用伪随机数来取代均匀随机数 U，如果取到的伪随机数落在 $(t_{i-1}, t_i]$ 中，则取 $F_i^{-1}(\frac{U-t_{i-1}}{p_i})$，这样得到的数就是 $\sum_{i=1}^{n} p_i F_i$ 混合分布随机数.

这个方法常用在排队论中．在那里的典型情形是混合指数分布(有的书上称为超指数分布)，即 $F_i(x) = 1 - e^{-l_i x}$，此时简单地有 $F_i^{-1}(y) = -\frac{1}{\lambda_i}\ln(1-y)$，于是计算变得非常简单而有效（当然也可利用命题 2.2 通过反函数来得到混合指数分布随机数．但是计算量会增加很多，因为这个反函数并不简单).

1.6 Von Neuman 取舍原则

Von Neuman 取舍原则：

假定我们要生成密度为 $p(x)$ 的随机数．为此取一个参考分布密度 $p_0(x)$，使它满足:

(1) $p_0(x)$ 随机数容易生成，例如 $p_0(x)$ 为正态密度，均匀密度，指数密度，及它们的混合密度；

(2) $p_0(x)$ 与 $p(x)$ 的取值范围差不多，且存在 C，使 $p(x) \le C \cdot p_0(x)$.

那么，我们有

命题 2．8 设随机变量 h 具有密度 $p_0(x)$，而随机变量 $U \sim U[0,1]$ 且与 h 独立，则

$$P(h \le x \mid \frac{p(h)}{Cp_0(h)} \ge U) = \int_{-\infty}^{x} p(v)dv .$$

证明 对 h 的取值用推广了的全概率公式（$P(A) = \int P(A \mid h = x)g(x)dx$），得到

$$左 = \frac{P(h \le x, \frac{p(h)}{Cp_0(h)} \ge U)}{P(\frac{p(h)}{Cp_0(h)} \ge U)} = \frac{\int_{-\infty}^{x} P(U \le \frac{p(y)}{Cp_0(y)})p_0(y)dy}{\int_{-\infty}^{\infty} P(U \le \frac{p(y)}{Cp_0(y)})p_0(y)dy} = \frac{\int_{-\infty}^{x} \frac{1}{C} p(y)dy}{\int_{-\infty}^{\infty} \frac{1}{C} p(y)dy} = 右 .$$?

取舍原则(Rejection Principle)的具体作法是:

(1) 独立地生成 n 个独立的 $p_0(x)$ 随机数 $h_1, \cdots, h_n$ 与 n 个与之独立的 $U[0,1]$ 随机数 $U_1, \cdots, U_n$.

(2) 对于$i=1,2,\cdots$，如果有$\dfrac{p(h_i)}{Cp_0(h_i)}\ge U_i$，就保留$h_i$，否则就舍弃$h_i$。

由命题 2.8，所有这样保留下来的那些h_i们就成为一系列独立的$p(x)$随机数(当然个数会比n小很多). 这种取舍方法称为 Von Neuman 取舍原则.

取舍原则可以改良为：

如果$p(x)=g\cdot h(x)$，只要存在C，使$h(x)\le C\cdot p_0(x)$，那么我们可以在取舍原则中用$h(x)$代替$p(x)$，得到$p(x)$随机数． 具体为：独立地生成n个独立的$p_0(x)$随机数$h_1,\cdots,h_n$与n个与之独立的独立$U[0,1]$随机数$U_1,\cdots,U_n$，如果$\dfrac{h(h_i)}{Cp_0(h_i)}\ge U_i$，则保留$h_i$，否则舍弃$h_i$，那么所有保留下的是相互独立的$p(x)$随机数.

证明　只需注意到这时有　$p(x)\le g\cdot C\cdot p_0(x)$，并且$\dfrac{p(x)}{g\cdot Cp_0(x)}=\dfrac{h(x)}{Cp_0(x)}$即可．

注１　一般地g需通过$\dfrac{1}{g}=\int h(x)dx$计算，其中的积分不易计算． 但是上面的事实说明不必计算g， 即可以忽视这个常数因子．这就使取舍原则变得非常好用．取舍原则的优点是简单易行，可以适用于非常复杂的分布.

注２ 如果$p(x)$只在有限区间$[a,b]$上不等于零，而且有界，那么$p_0(x)$就可取均匀分布 $U[a,b]$； 如果$p(x)$只在右半直线不等于零，那么指数分布就可以是$p_0(x)$的一个选择; 如果$p(x)$在实直线上不等于零，且分布密度的尾部不大， 则正态分布就可以是$p_0(x)$的一个选择；如果$p(x)$在实直线上不等于零，且分布密度的尾部较大（重尾分布）， 则t分布就可以是$p_0(x)$的一个选择； 如果$p(x)$具有多个峰，则混合正态分布或混合指数分布就可以是$p_0(x)$的一个选择. 可见适当精心地选取$p_0(x)$是使计算省时的关键.

注３　原则上取舍原则也适用于离散分布和多维密度，但是在多维密度的情形，$p_0(x)$的选取并不容易．

注４　样本生成的一个重要应用，就是对于函数的一些积分，用 Monte Carlo 方法(随机模拟算法)给出估计． 直观地可以猜测，采用的随机数的密度$p_0(x)$的形状与被积函数越像，则估计的方差会越小，即效果越好. 这种取样法称为重要度采样 (Impotance Samling)(确切的定义与方差的最小性证明可参见第 8 章)．

1. 7　Gamma 随机数与 Beta 随机数的生成

在$a \geq 1$时，设$[a]$是正数a的整部，可以证明$\Gamma(a, l)$分布的密度与$\Gamma([a], l)$分布的密度的比例是有界函数．由此可以设计一个由指数分布随机数，通过 Von Neuman 取舍原则得到$\Gamma(a, l)$分布的随机数的方法．

在$a, b \geq 1$时，设h_1, h_2分别为独立的$\Gamma(a,1), \Gamma(b,1)$随机数，则可以证明$V = \dfrac{h_1}{h_1 + h_2}$为$B(a, b)$随机数．

关于$0 < a < 1$时$G(a, l)$分布的取样方法，可以参见文献 [LK].

2. 多维随机数

2. 1　连续型多维随机数

对于已知的分布密度，我们可以利用条件密度，把生成多维随机数归结为生成一系列一维随机数．设随机向量$(X_1, \cdots, X_d)$的密度为$f(x_1, x_2, \cdots, x_d)$．那么，我们有表达式

$$f(x_1, x_2, \cdots, x_d) = f_{X_1}(x_1) f(x_2 \mid x_1) \cdots f(x_d \mid x_1, \cdots, x_{d-1}),$$

其中f_{X_1}为X_1的边缘密度，$f(x_k \mid x_1, \cdots, x_{k-1})$为在已知$X_1 = x_1, \cdots, X_{k-1} = x_{k-1}$条件下，$X_k$的条件密度．于是可以先取一个$f_{X_1}$随机数$x_1$；然后，在$x_1$固定的情形下，生成一个$f(\cdot \mid x_1)$随机数$x_2$；再在$x_1, x_2$固定的情形下，生成一个$f(\cdot \mid x_1, x_2)$随机数$x_3$；… 最后，在$x_1, \cdots, x_{d-1}$固定的情形下，生成一个$f(\cdot \mid x_1, \cdots, x_{d-1})$随机数$x_d$．这样得到的$(x_1, \cdots, x_d)$就是随机向量$(X_1, \cdots, X_d)$的一个随机数．当然在生成各个条件密度的随机数时，仍然可以使用 Von Neuman 取舍原则．　注意在用取舍原则于一维分布取样时，可以忽略一个因子．

然而，在d较大时，更为常用的是 Gibbs 取样法 (Gibbs Sampler)，它是一种基本的动态 Monte Carlo 方法，即 Markov 链 Monte Carlo 方法，在本书第 8 章中我们将介绍这种方法．

在很多实际情形中，多维密度常常并不知道，更无法知道各个条件密度．这时最自然而粗糙的想法是用条件频率来近似条件密度．

2. 2　离散型多维随机数

只要用概率函数$p(x_1, \cdots, x_d) \overset{\Delta}{=} P(X_1 = x_1, \cdots, X_d = x_d)$代替密度函数，2. 1 段中的办法就自然适用．

2. 3　多维正态随机数

设 $(X_1,\cdots,X_d)$ 服从 $N(\vec{m},\Sigma)$，其密度函数为 $p(\dot{x})=\frac{1}{(2p)^{\frac{d}{2}}|S|^{\frac{1}{2}}}e^{-\frac{1}{2}(\dot{x}-\dot{m})^T S^{-1}(\dot{x}-\dot{m})}$，

其中 $S=(s_{ij})_{i,j\le d}$ 为对称正定矩阵. 由线性代数我们知道，可以找到下三角形矩阵 C，

$(c_{ij}=0,\forall i<j)$，使 $\Sigma=CC^T$. 作独立 $N(0,1)$ 随机数 $x_1,\cdots,x_d$. 令 $\dot{x}\overset{\Delta}{=}(x_1,\cdots,x_d)^T$，则

$C\vec{x}+\vec{m}\sim N(\vec{m},\Sigma)$. 可见 $C\vec{x}+\vec{m}$ 为 $N(\vec{m},\Sigma)$ 随机数. 于是求多维正态随机数的快速生成问题，就归结为快速计算下三角形矩阵 C 的问题.

2.4 多维 Beta 随机数 (Dirichlet 随机数) 的生成

在 $a_1,\cdots,a_k\ge 1$ 时，设 $h_1,\cdots,h_k$ 分别为独立的 $\Gamma(a_1,1),\cdots,\Gamma(a_k,1)-$ 随机数，则可以证明

$$\left(\frac{h_1}{h_1+\cdots+h_k},\cdots,\frac{h_k}{h_1+\cdots+h_k}\right)$$

就是一个 Dirichlet $(a_1,\cdots,a_k)$ 随机数.

3．附录 一 用 Matalab 生成随机数

3．1 Matlab 语言的简单提示

Matlab 提供了强大功能的软件包，其用法可以参阅 Matlab 6.1(或 Matlab 6.5) 的 Help．为了使用方便，我们在本段列出一些基本的记号与最常见的语句．

一．Matlab 中的矩阵:

1．矩阵 A 的分量用方括号括起来．同行的分量用逗号或空格分开， 不同行的分量用分号或回车(Enter)分开． 空的方括号[ ]表示没有分量的矩阵，即空集．

2．A(3,6)表示矩阵 A 的(3,6)分量．反之， 给出了它的值也可以调用 A (在 A 的行列不足 3×4 时将自动增加行列， 即除了在该分量处赋要求的值以外，将其它分量增补为 0)．

3．矩阵运算除＋， －外， 还有：*（乘）， ' (转置)， ^ (幂)，

\ (左除， $A\backslash B$ 即 $A^{-1}B$)， / (右除，A/B 即 BA^{-1})．

常数视其情形可以看成任意阶的矩阵，如 A+a (A 的所有分量都加常数 a)．

4．可以用冒号对矩阵进行剪裁， 如 A[3,:]表示取 A 的第 3 行， A[:,3]表示取 A 的第 3 列，A[1:3,:] 表示取 A 的 1,2,3 行， A[1:2:3,:] 表示取 A 的 1,3 两行．又用 A[3,:]=[ ]表示删除 A 的第 3 行．

5．两个矩阵可以拼接成一个．

6．特殊的矩阵（在相应的语句后加 (m,n) 表示其阶）有：zeros(0 矩阵)，ones（分量全 1 的矩阵），eyes（对角为 1，其它为 0 的矩阵），rand ([0，1]均匀随机数的矩阵)等．

二 Matlab 中的数组

1．与矩阵的差别是数组不加括号．用冒号表示等差数组，即 $m:d:n$ 表示公差为 d，

起始于m，终止于n的等差数组，而将$m:1:n$简记为$m:n$

2．数组间在运算符号前加一个小点表示按分量运算，有：.*（乘），.^（幂），./ (除（即右除))， .\ (反除（即左除)，即除的倒数)．

3．几个数组可以用方括号拼接成矩阵．

4．单引号括起来的字符串可以作为数组．

又如语句$linspace(a,b,n)$生成起始于a，终止于b的n个等差数组．

三．逻辑关系

1．一个关系的逻辑值只取两个值：0（表示"错误"），或 1（表示"正确"）．所以，关系 x==0 的逻辑值，就是 x=0 的示性函数．

2．主要的逻辑关系与逻辑运算有：

==（逻辑相等)），>=（不小于），<=（不大于）， ~=（不等于）， &（与），|（或），~（非）．

四．语句，函数与操作命令

1．一般的语句的形式是

变量=表达式

(如用分号结束，则表示只运算不显示．如用回车结束则显示它的计算结果．如只有表达式，则自动生成变量 ans，并显示 ans =)．

2．两个语句间用逗号或分号．

3．特殊变量有 pi (圆周率p), i 或 j (虚数单位)，Inf （无穷大），NaN （不定值：0／0）， eps (主机中的最小的浮点数， 有时用它代替０可以防止溢出，即用表达式(x==0)*eps 代替 0， 它表示在 x=0 时用 eps 代替 x)．

4．函数的变量置于圆括号之中．常见的函数除对数，指数，三角函数外，还有：

atan（反正切）, abs（绝对值）, sqrt（开方）, round(4 舍 5 入取整), floor(负无穷方向取整), ceil(正无穷方向取整), fix(0 方向取整), imag（虚部）, rats(最近有理数)

length（数组的长度）, prod（乘）, sort(数组的分量按递增排序)，trace(求迹)，triu (矩阵取上三角), tril (矩阵取下三角), size（矩阵的大小）, norm（模）, inv（取逆）, eig（求特征值）, poly(求特征多项式), expm(求矩阵的指数函数), cond(条件的数目), lu(矩阵的 LU 分解), gr(矩阵的正交分解), svd(矩阵的奇异值分解)，

feval（赋值语句，如 feval('atan',x) 等价于 atan(x)．

5．用于循环的常见语句有：

pause（暂停），

all, any,

for, end

while, end,

if, elseif, else, end

switch, case, otherwise, end

break(跳出循环) ．

6．操作命令（很多与 Unix 命令相同）：

help, lookfor, format short（显示到小数点后４位）, whos（显示当前工作区的变量）, disp(x)(显示 x 的内容), clear, save, load, diary filename（建立新文件）, diary off（停止记录）, path, dir, type, delete, quit 等．

７．画图命令有：

plot(x,y)**（一个图）**, plot(x,y1,x,y2)**（两个图）**, **或** fplot('**图形名**',[xmin xmax ymin ymax])
或在图上加：hold on, **第２个图的语句**, hold off

线图： -， ：(**点线**), -. (**虚点线**)， ——(**波折**)

点图： ., +, *

颜色： y**（黄）**，r**（红）**，g**（绿）**，b**（兰）**，w**（白）**，k(**黑**)，m(**紫**)，c(**青** cyan)

例如：plot(x,y1,'b:',x,y2,'c-')

控制坐标范围用：axis auto, axis off

在图上标字：用 title('**字串**'), **或用** gtext('**字串**')**再用鼠标点**

分区画多个图：用 subplot(m,n,k) (**在**$m\times n$**个区中的第** k **个区中画图）**

曲面：[X,Y] ＝meshgrid(x,y)**；（把**(x，y)**改写为数组** [X,Y]**，以便用于表达式中）；**

Z＝(X,Y)**的表达式**; mesh(X,Y,Z)

空间曲线图例如：plot3(sin(t),cos(t),t)

等高线图：contour(X,Y,Z,m) (m **条等高线**)。

３．２ Matlab 生成随机数的语句

第一种方法是用 random **语句，其一般形式为**

y = random('**分布的英文名**',A1,A2,A3,m,n)，

表示生成m**行**n**列的**$m\times n$**个参数为**(A_1, A_2, A_3)**的该分布的随机数。例如**:

(1) R = random('Normal',0,２,2,4): **生成期望为** 0,**标准差为** 1 **的**(2 **行** 4 **列**)2×4 **个正态随机数**

(2) R = random('Poisson',1:6,1,6): **依次生成参数为** 1 **到** 6 **的**(1 **行** 6 **列**)6 **个** Poisson **随机数**

第二种方法是针对特殊的分布的语句：

一． 几何分布随机数 （下面的 P，m **都可以是矩阵）**

R = geornd(P) **（生成参数为** P **的几何随机数）**

R = geornd(P,m) **（生成参数为** P **的１**$\times m$ **个几何随机数）**

R = geornd(P,m,n) **（生成参数为** P **的**m**行**n**列的**$m\times n$**个几何随机数）**

例如

(1) R = geornd(1 ./ 2.^(1:6)) (**生成参数依次为** 1/2,1/2^2,**到** 1/2^6 **的** 6 **个几何随机数**)

(2) R = geornd(0.01,[1 5]) (**生成参数为** 0.01 **的（１行５列）**5 **个几何随机数**).

二．Beta **分布随机数**

R = betarnd(A,B) **（生成参数为** A,B **的** Beta **随机数）**

R = betarnd(A,B,m) **（生成１**$\times m$ **个数为** A,B **的** Beta **随机数）**

R = betarnd(A,B,m,n) **（生成**m**行**n**列的**$m\times n$**个数为** A,B **的** Beta **随机数）**.

三．正态随机数

R = normrnd(MU，SIGMA) **（生成均值为** MU，**标准差为** SIGMA **的正态随机数）**

R = normrnd(MU，SIGMA,m) **（生成１**$\times m$ **个正态随机数）**

R = normrnd(MU，SIGMA,m,n) **（生成**m**行**n**列的**$m\times n$**个正态随机数）**

例如

(1) R = normrnd(0,1,[1 5]) **生成** 5 **个正态**(0,1)**随机数**

(2) R = normrnd([1 2 3;4 5 6],0.1,2,3) **生成期望依次为**[1,2,3;4,5,6],**方差为** 0.1 **的** 6 **个正态随**

机数．

四．二项随机数：类似地有

R = binornd(N,P)   R = binornd(N,P,m)   R = binornd(N,p,m,n)

例如

n = 10:10:60; r1 = binornd(n,1./n) **或** r2 = binornd(n,1./n,[1 6]) **（都生成参数分别为** $(10,\frac{1}{10}),\cdots,(60,\frac{1}{60})$ **的 6 个二项随机数．**

五．自由度为 V **的** c^2 **随机数：**

R = chi2rnd(V)   R = chi2rnd(V,m)   R = chi2rnd(V,m,n)

六．期望为 MU **的指数随机数（即** $Exp_{\frac{1}{MU}}$ **随机数）：**

R = exprnd(MU)   R = exprnd(MU,m)   R = exprnd(MU,m,n)

七．自由度为 V1，V2 **的** F **分布随机数：**

R = frnd(V1，V2)   R = frnd(V1，V2,m)   R = frnd(V1，V2,m,n)

八．$\Gamma(A,l)$ **随机数：**

R = gamrnd（A,lambda）   R = gamrnd（A,lambda,m)   R = gamrnd（A,lambda,m,n)

九．超几何分布随机数：

R = hygernd(N,K,M)   R = hygernd(N,K,M,m)   R = hygernd(N,K,M,m,n)

十．对数正态分布随机数

R = lognrnd(MU，SIGMA)   R = lognrnd(MU，SIGMA,m)   R = lognrnd(MU，SIGMA,m,n)

十一．负二项随机数：

R = nbinrnd(r,p)   R = nbinrnd(r,p,m)   R = nbinrnd(r,p,m,n)

十二．Poisson **随机数：**

R = poissrnd(lambda)   R = poissrnd(lambda,m)   R = poissrnd(lambda,m,n)

例如，以下 3 **种表达有相同的含义：**lambda = 2; R = poissrnd(lambda,1,10)
（或 R = poissrnd(lambda,[1 10]) **或** R = poissrnd(lambda(ones(1,10)))

十三．Rayleigh **随机数：**

R = raylrnd(B)   R = raylrnd(B,m)   R = raylrnd(B,m,n)

十四．V **个自由度的** t **分布的随机数：**

R = trnd(V)   R = trnd(V,m)   R = trnd(V,m,n)

十五．离散的均匀随机数：

R = unidrnd(N)　　　R = unidrnd(N,m)　　R = unidrnd(N,m,n)

十六．[A,B]**上均匀随机数**

R = unifrnd(A,B)　　R = unifrnd(A,B,m)　　R = unifrnd(A,B,m,n)

例如 unifrnd(0,1:6)**与** unifrnd(0,1:6,[1 6])**都依次生成**[0,1]**到**[0,6]**的６个均匀随机数．**:

十七．Weibull **随机数**

R = weibrnd(A,B)　　R = weibrnd(A,B,m)　　R = weibrnd(A,B,m,n)

习题 2

1. **完成命题２．５的证明**.
2. **证明**

$$p_0(x)=\frac{l\,a^{l}\,x^{l-1}}{(a^{l}+x^{l})^2}$$

是一个分布密度. **设**U**是一个均匀随机数**, **取**$h=a\left(\frac{U}{1-U}\right)^{\frac{1}{l}}$, **证明它是一个** $p_0(x)$ **随机数**.

3. **设**h_1, h_2 **分别为独立的**$\Gamma(a,1),\Gamma(b,1)$ **随机数**, **证明** $V=\frac{h_1}{h_1+h_2}$ **为** $B(a,b)$ **随机数**.

4. **移位指数分布的密度为** $p(x)=\lambda e^{-\lambda(x-a)}I_{[a,\infty)}(x)$。 **如何得到它的样本 ？**

５．若随机变量$x \sim N(0,1)$**，** $h \sim c^2(n)$ **且与**x **独立，　则** $\frac{x}{\sqrt{\frac{h}{n}}}$ **服从** $t(n)$ **分布（其密度为**

$t(n,x)=\frac{G(\frac{n+1}{2})}{\sqrt{np}\,G(\frac{n}{2})}(1+\frac{x^2}{n})^{-\frac{n+1}{2}}$**）．如何构造**$t(n)$**随机数 ？**

6. **如何得到**$\Gamma(n,l), B(n,m)$**随机数 ？**

7. Weibull **分布的密度的另一个形式为** $f(x)=\frac{m}{a}(\frac{x}{a})^{m-1}e^{-(\frac{x}{a})^m}I_{[0,\infty)}(x)$ **．如何得到其随机数 ？**

8．证明若x **为** $\exp_l$ **随机数，　则**$h=ae^{\frac{lx}{r}}$ **为** Pareto **随机数．**

9．若x **为** Weibull **分布** $W(l,,a)$ **随机数，证明**$h=-b\ln(l\cdot x^{a})+m$ **为极值分布的随机数．**

10. 求密度为 $\dfrac{\Gamma(\frac{k+1}{2})}{\sqrt{kp}\,\Gamma(\frac{k}{2})}a^{-\frac{1}{2}}(1+\dfrac{x^2}{ak})^{-\frac{k+1}{2}}$ 的 $t(k,a)$ 分布的方差. 又问它的随机数应该如何得到 ?

(注意 $t(k,1)$ 分布就是 $t(k)$ 分布).

11. 给出下列随机数的取样程序: 二项分布 $B(10,\frac{1}{3})$, 参数为 $(5,\frac{1}{3})$ 的负二项分布, $(N,K,M)=(10,7,5)$ 的超几何分布, 各生成 50 个独立样本, 并求其均值, 方差, 直方图.

12. C. G. Park, T. Park 和 D.W. Shin 在 1996 的一个构造相关系数为给定值 r 的二维二值随机向量 (x_1,x_2) 的样本的方法如下: 令 $X_j \sim Poisson_{l_j}$ $(j=1,2,3)$, 且相互独立. 适当地选取 $l_j\,(j=1,2,3)$ 可以使 $x_i = I_{(0)}(X_i + X_3)(i=1,2)$ 满足要求. 为此只需证明

(1) $x_i \sim \begin{pmatrix} 0 & 1 \\ 1-p_i & p_i \end{pmatrix}$ $(i=1,2)$, 其中 $p_i = e^{-(l_i+l_3)}(i=1,2)$.

(2) $Cov(x_1,x_2) = p_1 p_2 (e^{l_3}-1)$.

最后, 由 p_1, p_2, r 解出 $l_j\,(j=1,2,3)$. 请对此设计得到 (x_1,x_2) 随机数的一个程序.

13. 设 $[a]$ 是正数 a 的整部, 证明 $\Gamma(a,l)$ 分布的密度与 $\Gamma([a],l)$ 分布的密度的比例是有界函数. 由此设计一个由指数分布随机数, 通过 Von Neuman 的取舍原则得到 $\Gamma(a,l)$ 分布的随机数的方法.

14. 设 $[a]$ 是正数 a 的整部, 证明 $B(a,b)$ 分布的密度与 $B([a],[b])$ 分布的密度的比例是有界函数. 由此设计一个由指数分布随机数, 通过 Von Neuman 的取舍原则得到 $B(a,b)$ 分布的随机数的方法.

15. 若 U 是 [0, 1] 均匀随机数, 则 $[\dfrac{lnU}{ln(1-p)}]$ ***+1*** (整数部分) 是参数为 p 的几何分布的随机数。

龚光鲁，钱敏平著 应用随机过程教程及其在算法与智能计算中的应用

清华大学出版社,2003

第 3 章 随机过程的一般概念与独立增量过程

1. 一般概念

1. 1 随机过程与有限维分布族

定义 3.1 设T 为$[0,\infty)$或（$-\infty$，∞）或R^d，依赖参数t（$t\in T$）的一族随机变量（或随机向量）$\{x_t\}$通称为随机过程，t称为时间．当T 为整数集或正整数集时，则一般称为随机序列．而当T 为二维（或更一般地，d 维）整数格点时，则称为随机场．

更明确地，随机过程x_t应该写成$x_t(w)$或x（t,w）．这里的w 代表做一次完整的试验，或者说，是一个基本事件．当w 取固定的值，例如w_0时，$x_t(w_0)$是t的函数，称为随机过程的一个轨道或一个"现实"，这个现实是由w_0导致的．

定义 3.2 随机过程或随机序列的概率特性，由它在任意有限个时刻$\{t_1,\cdots,t_n\}$ 上 $(x_{t_t},\cdots,x_{t_n})$的分布（称为有限维分布族）所确定．有限维分布族即:

$$\{F_{t_1,\cdots,t_n}(x_1,\cdots,x_n)\overset{\Delta}{=}P(x_{t_1}\le x_1,\cdots,x_{t_n}\le x_n):\forall n,\forall t_1,\cdots,t_n\in T,\forall x_1,\cdots,x_n\}.$$

定义 3.3 有限维分布都是 Gauss 分布的随机过程或随机序列，称为 Gauss 过程或 Gauss 序列．对于 Gauss 过程(或序列)$\{x_t:t\in T\}$，记$m(t)=Ex_t$, $s(t,s)=Cov(x_t,x_s)$，分别称为期望函数与协方差函数．$s(t,s)$是非负定对称函数，即

$$s(s,t)=s(t,s), s,t\in T\text{，矩阵}(s(t_i,t_j))_{i,j\le n}\text{为非负定矩阵（}\forall n,\forall t_1,\cdots,t_n\in T\text{）}.$$

于是$\{x_t:t\in T\}$的 Gauss 性就等价于：对任意有限个时刻$\{t_1,\cdots,t_n\}$，$(x_{t_t},\cdots,x_{t_n})$的矩母函数为

$$M(z_1,\cdots,z_n)=e^{m(t_1)z_1+\cdots+m(t_n)z_n+\frac{1}{2}\sum_{i,j\le n}s(t_i,t_j)z_iz_j}. \tag{3.1}$$

有时人们也用$R(t,s)\overset{\Delta}{=}E(\xi_t\xi_s)=\sigma(t,s)+m(t)m(s)$，称其为相关函数．可见 Gauss 过程的有限维分布族由期望函数与相关函数完全地确定了．

d 维随机过程$\overrightarrow{x_t}$是依赖于参数t的d 维随机向量族．其它概念与随机过程类似．

1. 2 独立增量过程

定义 3.4 称随机过程$\{x_t:t\ge 0\}$为独立增量过程，如果对于

$\forall n, \forall 0 \le t_0 < t_1 < \cdots < t_n$,起始随机变量及其后的增量$x_{t_0}, x_{t_1} - x_{t_0}, \cdots, x_{t_n} - x_{t_{n-1}}$是相互独立的随机变量组. 独立增量过程称为时齐的, 如果$x_{s+t} - x_s$的分布不依赖于s.

时间离散的独立增量过程, 就是独立随机变量的部分和. 而时齐的独立增量过程则是独立同分布随机变量部分和的时间连续情形.

记随机过程x_t的特征函数为 $\Phi(a,t) = Ee^{iax_t}$, 那么我们有

命题3.5 若$\xi_0 = 0$, 则独立增量过程x_t为时齐的必要充分条件为:其特征函数有可乘性, 即

$$\Phi(a,t+s) = \Phi(a,t)\Phi(a,s). \tag{3.2}$$

证明: 必要性显然.我们证明充分性.由独立增量性, 我们有下面的命题

$$\begin{aligned}&\Phi(a,t)\Phi(a,s) = \Phi(a,t+s)\\&= Ee^{iax_{s+t}} = Ee^{ia(x_{s+t}-x_s)}Ee^{iax_s} = Ee^{ia(x_{s+t}-x_s)}\Phi(a,s)\end{aligned},$$

由此得到

$$Ee^{ia(x_{s+t}-x_s)} = \Phi(a,t) = Ee^{iax_t},$$

这正说明了过程的时齐性.

独立增量过程具有以下的 Markov 性: 对于 $\forall s > s_1 > \cdots > s_m, \forall t, y, x, x_1, \cdots, x_m$有

$$P(x_{s+t} \le y \mid x_s = x, x_{s_1} = x_1, \cdots, x_{s_m} = x_m) = P(x_{s+t} \le y \mid x_s = x). \tag{3.3}$$

这个等式的推导将在本章第3节中在特殊情形(增量具有分布密度的情形)中给出. 等式(3.3)有非常明确的概率含义, 它说明了由这个随机过程所描写的随机现象具有以下特点: 在已知$x_s = x, x_{s_1} = x_1, \cdots, x_{s_m} = x_m$条件下, 随机变量$x_{s+t}$的条件分布函数只与$x_s = x$有关, 而与随机向量$(x_{s_1}, \cdots, x_{s_m})$的取值无关. 如果把$s$看成"现在", $x_s = x$看成现在的取值, 把$s+t$看成"将来", 小于s的时刻看成"过去", 那么这正是表达了: 对于独立增量过程, 在已知过去与现在的条件下, 将来的条件分布只与现在的取值有关, 而与过去的取值无关. 这种"忘记过去" 的性质, 称为无后效性或 Markov 性.

时齐的独立增量过程x_t具有非常特殊形式的特征函数: $\exists y(l)$,使$Ee^{ilx_t} = e^{-ty(l)}$.

类似地, 人们常遇到d维独立增量过程.

2 Poisson 过程与复合 Poisson 过程

2.1 事故申报次数的概率模型与 Poisson 过程

例3.6 (保险公司理赔次数) 设在时间间隔$(0,t]$中某保险公司收到的某类保险的理赔次数为N_t, 那么它是一个只取非负整值的随机过程.从长期经验的积累, 人们概括出以下

的初步近似性质：

(1) **在不同的时间区段内的理赔次数是彼此独立的随机变量**;

(2) **在同样长的时间区段内的理赔次数的概率规律是一样的**;

(3) $N_0=0$，**在有限时间区段内理赔次数是有限的，而且在非常短的时间区段** h **内的理赔次数超过** 2 **的概率是** h **的高价无穷小** $o(h)$.

(1)， (2) **说明了** N_t **是时齐的独立增量过程．而性质(3)称为普通性．**

我们来推导满足普通性的非负整值的时齐独立增量过程的概率分布．令 $p_i(t)=P(N_t=i)$，**我们有**

$$p_0(t+s)=P(N_{t+s}=0)=P(N_t=0,N_{t+s}-N_t=0)$$

$$=P(N_t=0)P(N_{t+s}-N_t=0)=p_0(t)p_0(s)\text{.}$$

解这个函数方程，可知存在 $l>0$，**使** $p_0(t)=e^{-lt}$． **再则，由性质(3)有**

$$P(N_{t+h}-N_t\ge 2)=o(h)\text{，}$$

利用时齐性及独立增量性，由全概率公式我们得到

$$p_{k+1}(t+h)=P(N_{t+h}=k+1)$$

$$=P(N_t=k,N_{t+h}-N_t=1)+P(N_t=k+1,N_{t+h}-N_t=0)+o(h)$$

$$=p_k(t)p_1(h)+p_{k+1}(t)p_0(h)+o(h)$$

$$=p_k(t)(1-p_0(h)-o(h))+p_{k+1}(t)e^{-lh}+o(h)$$

$$=p_k(t)(1-e^{-lh})+p_{k+1}(t)e^{-lh}+o(h)\text{.}$$

于是

$$\frac{p_{k+1}(t+h)-p_{k+1}(t)}{h}=p_k(t)\frac{1-e^{-lh}}{h}+p_{k+1}(t)\frac{e^{-lh}-1}{h}\text{.}$$

令 $h\to 0$，**便得无穷常微分方程组**:

$$p'_{k+1}(t)=\lambda\cdot p_k(t)-\lambda\cdot p_{k+1}(t) \qquad (3．4)$$

显见，对于 $k\ge 1$ **还应该满足初值条件：**

$$p_k(0)=0,\ p_0(0)=1.$$

下面我们用矩母函数方法来推导 $p_k(t)$ **的明显表达式．令**

$$M(t,z)\overset{\Delta}{=}Ez^{zN_t}=\sum_{k=0}^{\infty}z^k p_k(t)\text{.}$$

那么

$$\frac{\partial M(t,z)}{\partial t} = p'_0(t) + \sum_{k=0}^{\infty} p'_{k+1}(t) z^{k+1} = -l e^{-lt} + l\sum_{k=0}^{\infty}(p_k(t) - p_{k+1}(t)) z^{k+1}$$

$$= -l p_0(t) + l \cdot zM(t,z) - l(M(t,z) - p_0(t)) = l(z-1)M(t,z).$$

实质上这是一个以 z 为参数的关于自变量 t 的常系数常微分方程，且满足初始条件 $M(0,z) = p_0(0) = 1$. 易见此方程的解为

$$M(t,z) = e^{l(z-1)t} = \sum_{k=0}^{\infty} \frac{(lt)^k}{k!} e^{-lt} z^k .$$

按矩母函数的定义，由此得到表达式

$$p_k(t) = \frac{(lt)^k}{k!} e^{-lt} .$$

也就是说 $N_t \sim Poisson_{lt}$. 随机过程 N_t 称为 Poisson 过程.

由 Poisson 分布的性质推出：$EN_t = Var(N_t) = \lambda \cdot t$，由此我们便得到参数 l 的概率含义：$l = \frac{EN_t}{t} = \frac{Var(N_t)}{t}$，即 l 是单位时间的平均理赔次数，称为此Poisson过程的强度，同时它也代表单位时间理赔次数的方差.

[注 1] 方程(３．４)也可以用数学归纳法通过常微分方程中的常数变异法直接求解.

[注 2] Poisson 过程是用以描写一切”罕见事件”发生的概率规律的数学模型.

定义３．７ 时齐的独立增量过程 N_t 称为强度为 l 的 Poisson 过程，如果它满足

$$N_0 = 0, N_t \sim Poisson_{lt} .$$

Poisson 过程的联合分布为

$$t_1 < \cdots < t_m, n_1 < \cdots n_m, P(N_{t_1} = n_1, \cdots, N_{t_m} = n_m) = p_{n_1}(t_1) p_{n_2 - n_1}(t_2 - t_1) \cdots p_{n_m - n_{m-1}}(t_m - t_{m-1})$$

2. 2 Poisson 过程与指数流的关系

把理赔时刻看成随机到达的 ”点”，那么随着时间的发展，就出现一系列随机的点，记这些时刻为 $0 < t_1 < \cdots < t_m < \cdots \to \infty$，即 t_k 为第 k 个理赔发生的时刻. 这个概念可以抽象为下述定义

定义３．８ 如果随机序列 $\{t_k\}$ 满足：$0 = t_0 < t_1 < \cdots < t_m < \cdots \to \infty$，则称之为一个事件流，简称为流. 记 $T_k = t_k - t_{k-1}$. 如果 $T_k \sim \exp_l$，而且 $\{T_k\}$ 独立同分布，那么，我们称这个事件流 $0 = t_0 < t_1 < \cdots < t_m < \cdots \to \infty$ 为强度为 l 的指数流. 又由于随机序列

$\{T_k\}$ 与随机序列 $\{t_n)$ 唯一地相互确定，所以，有时我们也称 $\{T_k\}$ 为指数流.

对于指数流 $\{t_k)$ 而言，在时间段 $(0,t]$ 中出现的 τ_k 的个数，记为 $N_t = \sup\{k:\tau_k \le t\}$，是一个 Poisson 过程，$N_t$ 称为指数流的计数过程. 我们把这写成一般的结论.

定理3．9 对于取非负整值的随机过程 N_t， 令 $t_k = \inf\{t: N_t = k\}$ （它等价于 $N_t = \sup\{k: t_k \le t\}$)，那么下面诸事实彼此等价:

(1) $\{t_k\}$ 是强度为 l 的指数流;

(2) $\forall n, (t_1,\cdots,t_n)$ 的分布密度为 (记号 I_A 表示 A 的示性函数)

$$g(s_1,\cdots,s_n) = \lambda^n e^{-s_n} I_{\{0<s_1<\cdots<s_n\}}; \qquad (3.\ 4)$$

(3) N_t 是强度为 l 的 Poisson 过程.

证明 首先注意 $\{N_t \ge k\} = \{t_k \le t\}$.

$(1)\Rightarrow(2)$ 只要注意

$$P(t_1 \le s_1,\cdots,t_n \le s_n) = \int\cdots\int_{\substack{t_1,\cdots,t_n>0 \\ t_1\le s_1 \\ t_1+t_2\le s_2 \\ \cdots \\ t_1+\cdots+t_n\le s_n}} l^n e^{-l(t_1+\cdots+t_n)} dt_1\cdots dt_n = \int\cdots\int_{\substack{y_n>\cdots>y_1>0 \\ y_1\le s_1 \\ y_2\le s_2 \\ \cdots \\ y_n\le s_n}} l^n e^{-l y_n} dy_1\cdots dy_n$$

$$= \int\cdots\int_{\substack{y_1\le s_1 \\ y_2\le s_2 \\ \cdots \\ y_n\le s_n}} l^n e^{-l y_n} I_{0<y_1<\cdots<y_n} dy_1\cdots dy_n .$$

而 $(2)\Rightarrow(1)$ 得自 $\{t_n\}$ 与 $\{T_k\}$ 是一一对应的.

$(2)\Rightarrow(3)$ 先求 $\tau_k,(k\le n)$ 的密度(记成 $g_k(u)$). 由(2)用归纳法可得

$$g_k(u) = \int\cdots\int l^n e^{-l s_n} I_{0<s_1<\cdots<s_{k-1}<u<s_{k+1}<\cdots<s_n} ds_1\cdots ds_{k-1} ds_{k+1}\cdots ds_n$$

$$= \lambda^k e^{-\lambda u}\frac{u^{k-1}}{(k-1)!} I_{u>0} \quad (\Gamma(k,\lambda) \text{ 分布}). \qquad (3.\ 5)$$

再则，由(2)推出

$$P(N_s = m, N_{t+s} \ge m+k) = P(t_m \le s < t_{m+1} < \cdots < t_{m+k} \le s+t)$$

$$= \int\cdots\int_{\substack{s_m\le s \\ s<s_{m+1}<\cdots<s_{m+k}\le s+t}} l^{m+k} e^{-l s_{m+k}} I_{0<s_1<\cdots<s_{m+k}} ds_1\cdots ds_{m+k} .$$

采用变量替换 $y_{m+l}=s_{m+l}-s,(l\le k),u=y_{m+k}$ 后，可以很容易地算出右方的积分为 $l^m\frac{s^m}{m!}e^{-ls}\int_0^t g_k(u)du$，即我们得到了

$$P(N_s=m,N_{t+s}\ge m+k)=l^m\frac{s^m}{m!}e^{-ls}\int_0^t g_k(u)du.$$

进而有

$$P(N_{s+t}-N_s\ge k)=\sum_{m=0}^{\infty}P(N_s=m,N_{s+t}\ge m+k)$$

$$=\sum_{m=0}^{\infty}\lambda^m\frac{s^m}{m!}e^{-\lambda s}\int_0^t g_k(u)du=\int_0^t g_k(u)du.$$

并且

$$P(N_{s+t}-N_s=k)=P(N_{s+t}-N_s\ge k)-P(N_{s+t}-N_s\ge k+1)=\int_0^t(g_k(u)-g_{k+1}(u))du$$

$$=\int_0^t\{\frac{\lambda^k u^{k-1}}{(k-1)!}e^{-\lambda u}-\frac{\lambda^{k+1}u^k}{k!}e^{-\lambda u}]du=[\frac{\lambda^k u^k}{k!}e^{-\lambda u}]_0^t=\frac{(\lambda\cdot t)^k}{k!}e^{-\lambda t}.$$

特别地还有 $P(N_t=k)=P(N_t-N_0=k)=\frac{(\lambda\cdot t)^k}{k!}e^{-\lambda t}$. 由此可见

$$P(N_s=m,N_{t+s}-N_s\ge k)=P(N_s=m,N_{t+s}\ge m+k)=\lambda^m\frac{s^m}{m!}e^{-\lambda s}\int_0^t g_k(u)du$$

$$=P(N_t=k)\ P(N_{s+t}-N_s\ge k).$$

作为推论，我们得到 $P(N_s=m,N_{t+s}-N_s=k)=P(N_t=k)\ P(N_{s+t}-N_s=k)$. 这说明了 N_s 与 $N_{s+t}-N_s$.的独立性. 利用类似的推理, 对于 $\forall n,\forall 0<t_1<\cdots<t_n$, 可得

$$P(N_s=m,N_{s+t_1}-N_s=m_1,\cdots,N_{s+t_n}-N_{s+t_{n-1}}=m_n)$$

$$=P(N_s=m)P(N_{s+t_1}-N_s=m_1)\cdots P(N_{s+t_n}-N_{s+t_{n-1}}=m_n)$$

$$=\frac{(\lambda\cdot s)^m}{m!}e^{-\lambda s}\ \frac{(\lambda\cdot t_1)^{m_1}}{m_1!}e^{-\lambda t_1}\cdots\frac{(\lambda\cdot(t_n-t_{n-1}))^{m_n}}{m_n!}e^{-\lambda(t_n-t_{n-1})}.$$

这就证明了随机过程 x_t 的独立增量性与 Poisson 性.

$(3)\Rightarrow(2)$ 的证明

对于$\forall n, \forall 0 < s_1 < \cdots < s_n$，取充分小的$h_1, \cdots, h_n (< \max_{i \le n}(s_i - s_{i-1}), s_0 = 0)$

$$P(s_1 < t_1 < s_1 + h_1, \cdots, s_n < t_n < s_n + h_n)$$

$$= P(N_{s_1} = 0, N_{s_1+h_1} - N_{s_1} = 1, N_{s_2} - N_{s_1+h_1} = 0, \cdots, N_{t_n+h_n} - N_{t_n} = 1)$$

$$= P(N_{s_1} = 0)P(N_{h_1} = 1)P(N_{s_2-s_1-h_1} = 0)\cdots P(N_{h_n} = 1)$$

$$= e^{-\lambda s_1} \cdot \lambda \cdot h_1 e^{-\lambda h_1} \cdot e^{-\lambda(s_2-s_1-h_1)} e^{-\lambda s_2} \cdots \lambda \cdot h_n e^{-\lambda h_n} + o(h_1 \cdots h_n)$$

$$= \lambda^n h_1 \cdots h_n e^{-\lambda s_n} + o(h_1 \cdots h_n).$$

除以$h_1 \cdots h_n$后，令$h_1, \cdots, h_n \to 0$，便得到$(t_1, \cdots, t_n)$在约束条件$0 < s_1 < \cdots < s_n$下的分布密度为$\lambda^n e^{-\lambda s_n}$. (2)得以证明.

定理３．１０ (Poisson 过程的随机分流定理)

设N_t为强度为 l 的 Poisson 过程, 如果把其相应的指数流看成顾客流, 用与此指数流相互独立的概率 p ，把每个到达的顾客， 归入第一类, 而以概率$1-p$ 把他归入第二类. 对$i = 1,2$，记$N_t^{(i)}$为t前到达的第i类顾客数. 那么$\{N_t^{(1)} : t \ge 0\}$与$\{N_t^{(2)} : t \ge 0\}$分别为强度 pl 与$(1-p)l$ 的 Poisson 过程, 而且这两个过程相互独立. (这个性质称为 Poisson 过程的随机分流定理，也称为 Poisson 过程在随机选取下的不变性).

证明 由N_t是独立增量过程及归类的机制，可知$N_t^{(i)}$都是独立增量过程, 而且

$$P(N_{t+h}^{(i)} - N_t^{(i)} \ge 2) = o(h),\ P(N_{t+h}^{(i)} - N_t^{(i)} = 1) = \begin{cases} pl \cdot h + o(h), (i = 1) \\ (1-p)l \cdot h + o(h), (i = 2) \end{cases}.$$

所以它们都是 Poisson 过程. 下面我们先证明它们在同一个时刻的独立性: 由于

$$P(N_t^{(1)} = n, N_t^{(2)} = m \mid N_t = n+m) = C_{n+m}^n p^n (1-p)^m,$$

我们有

$$P(N_t^{(1)} = n, N_t^{(2)} = m) = C_{n+m}^n p^n (1-p)^m \cdot e^{-\lambda t} \frac{(\lambda \cdot t)^{n+m}}{(n+m)!}$$

$$= e^{-p\lambda} \frac{(p\lambda \cdot t)^n}{n!} \cdot e^{-(1-p)\lambda} \frac{((1-p)\lambda \cdot t)^m}{m!} = P(N_t^{(1)} = n)P(N_t^{(2)} = m).$$

这就证明了在固定的时刻t, $N_t^{(1)}$与$N_t^{(2)}$独立. 我们用类似而较为冗长的叙述，可以证明对于任意n, m, 及$t_1, \cdots, t_n; s_1, \cdots, s_m$，随机向量$(N_{t_1}^{(1)}, \cdots, N_{t_n}^{(1)})$与$(N_{s_1}^{(2)}, \cdots, N_{s_m}^{(2)})$的

独立性. 这就是说，随机过程$\{N_t^{(1)}: t\ge 0\}$ 与 $\{N_t^{(2)}: t\ge 0\}$是独立的.

［注］指数流与 Poisson 过程的离散时间版本

令T_k为独立同分布的几何分布随机序列，又$t_k = T_1+\cdots+T_k$，$\{N_n = k\}=\{\tau_k = n\}$.

记参数(k, p)的负二项分布为$NB(k;p)$，即

$$P(NB(k;p)=n)=C_{n-1}^{k-1}(1-p)^{n-k}p^k.$$

由简单的概率计算可得到$t_k \sim NB(k;p)$，从而$P(N_n = k)=P(NB(k;p)=n)$. 此处的N_n正是起到"离散时间的 Poisson 过程"的作用， 即它就是"离散时间情形的 Poisson 过程". 我们把它列表对比如下:

	流的间隔T_k	流的到达时刻t_k	计数过程N_t，或N_n
连续型:指数流	指数分布Exp_λ	$\Gamma(k, l)$分布	$\text{Poisson}_{\lambda t}$
离散型:几何流	几何分布	负二项$NB(k;p)$	$P(N_n = k)=P(NB(k;p)=n)$

2.3 与指数流有关的一些随机变量与分布

定理 3.11 若N_t为 Poisson 过程, 则在$N_t = n$的条件下, $(t_1,\cdots,t_n)$的条件分布密度为

$$f(s_1,\cdots,s_n)=\frac{n!}{t^n}I_{0<s_1<\cdots<s_n\le t}. \quad (3.6)$$

也就是说, 如果$h_1,\cdots,h_n$独立且服从 U[0,t], 而$h_{(1)},\cdots,h_{(n)}$为$h_1,\cdots,h_n$按次序大小重新排列而得的顺序随机变量: $h_{(1)}\le\cdots\le h_{(n)}$. 那么在$N_t = n$的条件下, $(t_1,\cdots,t_n)$的分布密度与$(h_{(1)},\cdots,h_{(n)})$的分布密度相同.

[注] 把上面的证明倒回去, 就可以发现此定理的结论反过来也是对的, 即: 如果一个取非负整值的跃度为 1 的非降随机过程N_t， 满足: $N_t \sim Poisson_{lt}$, 且在$N_t = n$的条件下, $(t_1,\cdots,t_n)$的条件分布密度为$f(s_1,\cdots,s_n)=\frac{n!}{t^n}I_{0<s_1<\cdots<s_n\le t}$, 其中$t_n$为$N_t$的第$n$次跳跃时刻, 那么$N_t$是 Poisson 过程.

我们还有下述相关的结论：

(1) 在$N_t = n$的条件下, t_n的条件分布密度为

$$g_n(s)\stackrel{\Delta}{=}\frac{ns^{n-1}}{t^n}I_{[0,t]}(s) \quad (3.7)$$

(2) $$P(\tau_n\le s, N_t = n)=\frac{(\lambda s)^n}{n!}e^{-\lambda t}I_{[0,t]}(s) \quad (3.8)$$

证明 对于$\forall n, \forall 0<s_1<\cdots<s_n$, 取充分小的$h_1,\cdots,h_n(<\max_{i\le n}(s_i - s_{i-1}), s_0 = 0)$,

就有

$$P(s_1<t_1<s_1+h_1,\cdots,s_n<t_n<s_n+h_n\mid N_t=n)$$

$$=\frac{P(s_1<t_1<s_1+h_1,\cdots,s_n<t_n<s_n+h_n,N_t=n)}{P(N_t=n)}$$

$$=\frac{P(N_{s_1}=0)P(N_{h_1}=1)P(N_{s_2-s_1-h_1}=0)\cdots P(N_{h_n}=1)P(N_{t-s_n-h_n}=0)}{\frac{(lt)^n}{n!}e^{-lt}}$$

$$=\frac{e^{-\lambda s_1}\cdot\lambda\cdot h_1e^{-\lambda\cdot h_1}\cdot e^{-\lambda(s_2-s_1-h_1)}\cdots\lambda\cdot h_ne^{-\lambda\cdot h_n}\cdot e^{-\lambda(t-s_n-h_n)}+o(h_1\cdots h_n)}{\frac{(\lambda\cdot t)^n}{n!}e^{-\lambda t}}$$

$$\to\frac{n!}{t^n},(h_1,\cdots,h_n\to 0)$$

定理 3.12 若 $s\le t,k\le n$，则在 $N_t=n$ 条件下，N_s 的条件分布与二项分布 $B(n,\frac{s}{t})$ 相同. 而 t_k 的条件分布则是

$$P(t_k\le s\mid N_t=n)=\sum_{j=k}^{n}C_n^j\left(\frac{s}{t}\right)^j\left(1-\frac{s}{t}\right)^{n-j}\qquad(3.9)$$

证明

$$P(t_k\le s\mid N_t=n)=\frac{\sum_{j=k}^{n}P(t_j\le s,t_{j+1}>s,N_t=n)}{P(N_t=n)}$$

$$=\frac{\sum_{j=k}^{n}P(N_s=j,N_t-N_s=n-j)}{\frac{(l\cdot t)^n}{n!}e^{-lt}}=\sum_{j=k}^{n}C_n^j(\frac{s}{t})^j(1-\frac{s}{t})^{n-j}\ .$$

推论3.13 若 $s\le t,k\le n$, 则 $E(t_k\mid N_t=n)=\frac{kt}{n+1}$.

证明 用次序统计量的结果：独立同分布的 $U[0,t]$ 的随机变量，按小至大的第 k 个次序随机变量的期望为 $\frac{kt}{n+1}$.

推论3.14 $E\tau_{N_t}=\frac{e^{-\lambda t}-1+\lambda t}{\lambda\cdot t}$

证明 $Et_{N_t}=E[E(t_{N_t}\mid N_t)]=\sum_{n=0}^{\infty}E(t_n\mid N_t=n)P(N_t=n)$

$$= \sum_{n=0}^{\infty} \frac{nt}{n+1} \frac{(\lambda \cdot t)^n}{n!} e^{-\lambda \cdot t} = \sum_{n=0}^{\infty} [(n+1)-1] \frac{1}{\lambda} \frac{(\lambda \cdot t)^{n+1}}{(n+1)!} e^{-\lambda \cdot t}$$

$$= \frac{e^{-\lambda \cdot t}}{\lambda} [(\lambda \cdot t e^{\lambda \cdot t} - (e^{\lambda \cdot t} - 1)] = \frac{e^{-\lambda \cdot t} - 1 + \lambda \cdot t}{\lambda} .$$

下面给出当前所用部件的寿命分布

把T_k**解释成某工作线上第**k**次被更新的部件的寿命． 假定它们都服从分布**$\exp_l$，**那么它们所对应的计数过程**N_t**就是** Poisson **过程**. **考虑当前时刻**t**所用的部件** (**注意当前时刻**t**所用的部件不是第**N_t**个**, **而是**N_t+1 **个，因为更新了**N_t**次后，起用的应该是第**N_t+1**个部件**) **的寿命**T_{N_t+1}. **这里**T_{N_t+1}**的随机性，不仅来自固定的部件的寿命**, **而且还来自**N_t**的随机性**, **所以不应该认为它服从指数分布**. **以下我们推导它的分布**.

因为$t_{N_t} \le t < t_{N_t} + T_{N_t+1}$, **所以**$T_{N_t+1} \le s$**就等价于**$t - t_{N_t} < T_{N_t+1} \le s$. **于是：**

(1) **当**$s < t$**时**, **由**$T_{N_t+1} \le s$**，我们有**$N_t \ge 1$**且**$t - s \le t_{N_t} < t$. **利用**$N_t = n$**与**T_{n+1}**之间的独立性及**t_n**与**T_{n+1}**的独立性**, **由**(3. **7**)**式所给出对条件密度函数**$g_n(u)$**导致**

$$\begin{aligned} P(T_{N_t+1} \le s) &= P(t - t_{N_t} < T_{N_t+1} \le s, t - s \le t_{N_t} < t) \\ &= E(E[I_{[t-t_{N_t}, s]}(T_{N_t+1}) I_{[t-s,t)}(t_{N_t}) \mid N_t, t_{N_t}]) \end{aligned}$$

$$= \sum_{n=1}^{\infty} \int E[I_{[t-\tau_n, s]}(T_{n+1}) I_{[t-s,t)}(\tau_n) \mid N_t = n, \tau_{N_t} = u) g_n(u) du \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

$$= \sum_{n=1}^{\infty} \int_{t-s}^{t} (e^{-\lambda(t-u)} - e^{-\lambda s}) \frac{n u^{n-1}}{t^n} du \frac{(\lambda \cdot t)^n}{n!} e^{-\lambda t}$$

$$= \int_{t-s}^{t} (e^{-\lambda(t-u)} - e^{-\lambda s}) \lambda \cdot du = 1 - e^{-\lambda s} - \lambda s e^{-\lambda s} .$$

(2) **当**$s \ge t$**时**, **我们有**$0 \le t_{N_t} \le t$, **并注意用**$t_0 = 0$, **类似地得到**

$$\begin{aligned} P(T_{N_t+1} \le s) &= P(t - t_{N_t} < T_{N_t+1} \le s, 0 \le t_{N_t} < t) \\ &= E(E[I_{[t-t_{N_t}, s]}(T_{N_t+1}) I_{[0,t)}(t_{N_t}) \mid N_t, t_{N_t}]) \end{aligned}$$

$$= \sum_{n=1}^{\infty} \int_{0}^{t} (e^{-\lambda(t-u)} - e^{-\lambda s}) \frac{n u^{n-1}}{t^n} du \frac{(\lambda \cdot t)^n}{n!} e^{-\lambda t} + P(t < T_1 \le s, N_t = 0)$$

$$= \int_{0}^{t} (e^{-\lambda(t-u)} - e^{-\lambda s}) \lambda \cdot du + (e^{-\lambda t} - e^{-\lambda s}) = 1 - \lambda t e^{-\lambda s} - e^{-\lambda s} .$$

综合(1)**与**(2)**，我们得到**

命题３．１５　当前时刻t所用的部件的寿命T_{N_t+1}的概率分布密度函数为

$$p_{T_{N_t+1}}(s)\overset{\Delta}{=}\lambda(1+\lambda\cdot s)e^{-\lambda s}I_{[0,t)}(s)+\lambda(1+\lambda\cdot t)e^{-\lambda s}I_{[t,\infty)}(s).$$

例３．１６　设第k个乘客到达公共汽车站的时刻t_k，服从指数流，则在$[0,t]$中所有乘客等待时间的和的数学期望为

$$E(\sum_{k=1}^{N_t}(t-t_k))=E\{E(\sum_{k=1}^{N_t}(t-t_k)\mid N_t)\}$$

$$=\sum_{n=0}^{\infty}P(N_t=n)E(\sum_{k=1}^{n}(t-t_k)\mid N_t=n)$$

$$=\sum_{n=0}^{\infty}P(N_t=n)(nt-\sum_{k=1}^{n}E(t_k\mid N_t=n)).$$

由定理３．１１可知上式右方第二项等于n个独立的$U[0,t]$随机变量和的期望，其值应为$\frac{nt}{2}$，于是

$$E(\sum_{k=1}^{N_t}(t-t_k))=\sum_{n=0}^{\infty}\frac{(lt)^n}{n!}e^{-lt}(nt-\frac{nt}{2})=\frac{lt^2}{2}.$$

2．4　**常见的推广**

１．非时齐的 Poisson 过程

对于独立增量过程N_t，如果存在可积正函数$l(t)$使 $N_t-N_s\sim\exp_{\int_s^t l(u)du}$，那么$N_t$就称为强度函数为$l(t)$的非时齐 Poisson　过程．而 Poisson 过程的强度函数为常数．

定理 3．１0’　(**时齐** Poisson **过程的非齐次分流定理)**

设N_t为强度为 l 的 Poisson **过程，把其相应的指数流看成顾客流，对任意时刻s，如有顾客到达，则以与此指数流相独立的概率$p(s)$ $(0<p(s)<1)$，确定该顾客归入第一类，而以概率$1-p(s)$确定该顾客归入第二类．记$N_t^{(i)}$为t前到达的第i类顾客数．那么$\{N_t^{(1)}:t\ge 0\}$与$\{N_t^{(2)}:t\ge 0\}$分别为强度 $l\int_0^t p(s)ds$与$l\int_0^t(1-p(s))ds$的非时齐** Poisson **过程，而且相互独立．**

(**证明与** Poisson **过程的分流定理相仿．只要注意**

$$P(N_{t+h}^{(i)}-N_t^{(i)}=1)=\begin{cases}p(t)\,l\cdot h+o(h),(i=1)\\(1-p(t))\,l\cdot h+o(h),(i=2)\end{cases}).$$

（这个定理还可以推广到分成多个类的情形）.

2．带随机调制的 Poisson 过程（二重 Poisson 过程）

设 $\{V_t\}$ 为随机过程，若在 $\{V_t\}$ 已知的条件下，非负整值随机过程 N_t 为强度为 $l(V_t)$ 的非时齐 Poisson 过程，则 N_t 称为带随机调制 $\{V_t\}$ 的 Poisson 过程，也称为二重 Poisson 过程．而更常见的是带随机线性自调制 $l \cdot N_t$ 的 Poisson 过程，也称为（0－记忆）自激点过程（参见第１７章第５节）.

非负随机变量也常称为一个随机时间． 随机时间 t 称为 $\{V_t\}$ -可知的，如果对于任意 t，事件 $\{t \le t\}$ 可由过程 $\{V_t\}$ 在 t 的信 息：$\{V_s : s \le t\}$ 所决定． $\{V_t\}$ -可知的随机时间又称为 $\{V_t\}$ -停时．对于带随机调制 $\{V_t\}$ 的 Poisson 过程 N_t，假定它所决定的流为 $\{t_n\}$，则 t_n 们也都是 $\{V_t\}$ -可 知的，也就是说，它们都是 $\{V_t\}$ -停时.

3 二维 Poisson 过程

设有两类不同的”事故”：事故 I 及事故 II. 记[0,t]内发生事故 I 及事故 II 的次数分别为 $\overrightarrow{N_t} = (N_t^{(1)}, N_t^{(2)})$，并设它们满足:

(1) 在不同的时间区段内事故申报数是独立的;

(2) 在同样长的时间区段内事故申报数的联合概率规律是一样的;

(3) $\vec{N}_0 = (0,0)$，又假定在有限时间区段内，两种事故申报数是有限的，而且在非常短的时间区段 Δt 内任意一种事故申报数超过 2 的概率相对于 Δt 为高价无穷小 $o(\Delta t)$ (即

$$P(N_{\Delta t}^{(1)} \ge 2 \text{或} N_{\Delta t}^{(2)} \ge 2) = o(\Delta t), \quad P(N_{\Delta t}^{(1)} = 1, N_{\Delta t}^{(2)} = 0) = \lambda_1 \Delta t,$$

$$P(N_{\Delta t}^{(1)} = 0, N_{\Delta t}^{(2)} = 1) = \lambda_2 \Delta t, P(N_{\Delta t}^{(1)} = 1, N_{\Delta t}^{(2)} = 1) = m\Delta t).$$

仿照 Poisson 过程情形，我们可以推导得 $\overrightarrow{N_t}$ 的矩母函数为

$$Ee^{z_1 N_t^{(1)} + z_2 N_t^{(2)}} = e^{-l_1(z_1-1) - l_2(z_2-1) - m(z_1 z_2 - 1)}.$$

由它的展开式就可以得到 $\overrightarrow{N_t}$ 的各种概率分布.

2. 5 复合 Poisson 过程

设 $\{h_n\}$ 为独立同分布序列，而 N_t 是一个与它独立的强度为 l 的 Poisson 过程．我们称随机过程 $V_t = h_1 + h_2 + \cdots + h_{N_t}$ 为 (强度为 l 的) 复合 Poisson 过程.

V_t 的特征函数为

$$F(a, t+s) = Ee^{iaV_{t+s}} = Ee^{ia(h_1 + \cdots + h_{N_{t+s}})}$$

$$=\sum_{n=0}^{\infty} E(e^{ia(h_1+\cdots+h_n)} \mid N_{t+s}=n)P(N_{t+s}=n\} = \sum_{n=0}^{\infty} Ee^{ia(h_1+\cdots+h_n)} \cdot \frac{(l\cdot(t+s))^n}{n!}e^{-l(t+s)}$$

$$=\sum_{n=0}^{\infty}\frac{(lEe^{iah_1}(t+s))^n}{n!}e^{-l(t+s)} = e^{l(Ee^{iah_1}-1)(t+s)}$$

由此可以看出

$$\Phi(a,t)=e^{l(Ee^{iah_1}-1)t}, \Phi(a,t+s)=\Phi(a,t)\Phi(a,s).$$

下面我们来验证它是时齐的独立增量性. 为了突出其实质, 我们只看最简单的情形， 即验证$V_t, V_{t+s}-V_s$相互独立. 对于任意有界函数f, g我们有

$$E[f(V_t)g(V_{t+s}-V_t)] = E[f(h_1+\cdots+h_{N_t})g(h_{N_t+1}+\cdots+h_{N_{t+s}})]$$

$$=\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}E([f(h_1+\cdots+h_l)g(h_{l+1}+\cdots+h_{l+k})] \mid N_t=l, N_{t+s}=l+k)P(N_t=l, N_{t+s}=l+k)$$

$$=\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}E[f(h_1+\cdots+h_l)g(h_{l+1}+\cdots+h_{l+k})]P(N_t=l, N_{t+s}-N_t=k)$$

$$=\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}Ef(h_1+\cdots+h_l)Eg(h_{l+1}+\cdots+h_{l+k})P(N_t=l)P(N_s=k)$$

$$=\sum_{l=\infty}^{\infty}\sum_{0k=0}Ef(h_1+\cdots+h_l)Eg(h_1+\cdots+h_k)P(N_t=l)P(N_s=k)$$

$$=[\sum_{k=0}^{\infty}Ef(h_1+\cdots+h_l)P(N_t=l)][\sum_{k=0}^{\infty}Eg(h_1+\cdots+h_l)P(N_s=k)]$$

$$=Ef(h_1+\cdots+h_{N_t})Eg(h_1+\cdots+h_{N_s})$$

$$=Ef(V_t)Eg(V_s)$$

由f, g的任意性可见$V_t, V_{t+s}-V_t$是相互独立的, 而且V_s与$V_{t+s}-V_t$同分布. 完全类似可以证明对于$0=t_0<\cdots<t_n, V_{t_1}-V_{t_0}, \cdots, V_{t_n}-V_{t_{n-1}}$也是相互独立的. 综上可知, 复合 Poisson 过程是时齐的独立增量过程, 因而它也是时齐的 Markov 过程. 在实际应用中, 时段$(0,t]$内, 保险公司对某项保险支付的累计理赔金额, 设备故障所需的累计维修费, 自然灾害所造成的累计损失, 股票市场的累计价格变动等等, 都可以用复合 Poisson 过程来近似地描述.

3 Brown 运动(Wiener 过程)及其函数

3．1 历史背景与物理模型

1827 年英国生物学家 Brown 在显微镜下, 观测悬浮在液面上的花粉，发现花粉微粒作着高度不规则的运动. 以后其他科学家发现了更多的类似现象，如空气中的烟雾的扩散等，但是一直找不出理想的模型来刻画此类现象. 至 19 世纪末，人们才搞清楚这种奇怪的现象是由于花粉（烟尘微粒）受到大量液体（空气）分子的无规则碰撞而造成的.

1905 年 Einstein 首次对此类现象作了理论上的量化分析. 他假定浸没在某种介质中的粒

子连续不断地受到周围介质中的分子的冲击，从物理的角度解释了这种现象.

以后，Ornstein 和 Uhlenbeck 等在物理上又进一步完善了这个想法.

在数学上严格地描述 Brown 发现的这种无规则运动，并把它纳入随机过程框架的是 Wiener. 他自 1918 年起系统地用随机过程来建立这种运动的数学模型. 这个随机过程也因此被称为 Wiener 过程. 又因为 Wiener 过程的背景源起于 Brown 的研究，所以也称为 Brown 运动. 实际上，早在 Einstein 提出理论解释之前，法国数学家 Bachellier 于 1900 年在研究债券市场时，就已经用类似的想法得到了 Brown 运动的直观模型，只是他的结果在经过半个世纪多以后才被人们重新发现.

至今，对于作为数学模型的 Brown 运动的理论研究，已经非常完善。 Brown 运动与 Poisson 过程一起，可以说是随机过程的两大基石. Brown 运动的应用十分宽广，几乎遍及自然科学和人文科学的所有领域. 有些 Nobel 经济学奖获得者的工作，是以 Brown 运动理论作为基础的.

物理模型的建立

按照 Einstein 的分析，Brown 运动表达了一个作随机运动的粒子在时间$[0,t]$上的位移$\{B_s: 0\le s\le t\}$. 不妨设$B_0=0$，Einstein 认为该粒子运动应满足以下的性质：

（1）粒子位移的各分量都相互独立. 所以，我们不妨只考虑其一个分量，仍记之为$\{B_t, t\ge 0\}$，且假定这个分量是时齐的独立增量过程;

（2）运动的统计规律对空间是对称的，从而$EB_t=0$;

（3）对固定的 $t>0$，B_t是一连续型随机变量，假定$g(t)\overset{D}{=}E(B_{t+h}-B_h)^2$存在,而且是 t 的连续函数.

下面对于固定的t，我们来推导B_t的分布密度的表达式. 记B_t的分布密度为$p(t,x)$，那么，由性质（2）及$B_0=0$可知

$$g(t)=E(B_t)^2=\int_{-\infty}^{\infty}x^2p(t,x)dx\text{.}$$

再由(3)得到

$$g(t+s)=E(B_{t+s})^2=E(B_{t+s}-B_t+B_t-B_0)^2$$

$$=E(B_{t+s}-B_t)^2+E(B_t-B_0)^2=g(s)+g(t)\text{.}$$

又因为$g(t)$是t的连续函数，所以$g(t)$一定是 t 的齐次线性函数，即

$$g(t)=Dt$$

(D 是一待定常数，它是单位时间内粒子平方位移的均值，称之为扩散常数，在分子运动学中，可知$D=\dfrac{RT}{Nf}$，其中 R 是由分子的特性所决定的一个普适常数，T 是绝对温度，N 是

Avogadro 常数，f 是摩擦系数).　在数学中为了方便常常简单地假定 $D=1$．于是我们得到

$$Var(B_t)=t$$

记 B_t 的特征函数为 $j(t,l)=Ee^{ilB_t}$　$(-\infty< l <\infty)$。 那么，利用条件（1）和（3），我们可得

$$\varphi(t+s,\lambda)-\varphi(t,\lambda)=Ee^{i\lambda B_{t+s}}-Ee^{i\lambda B_t}$$

$$\begin{aligned}
&=E(e^{ilB_t}(e^{il(B_{t+s}-B_t)}-1))=E(e^{il(B_t-B_0)}(e^{il(B_{t+s}-B_t)}-1))\\
&=E(e^{il(B_t-B_0)})E(e^{il(B_{t+s}-B_t)}-1)=E(e^{ilB_t})E(e^{il(B_s-B_0)}-1) \qquad \text{（E．1）}\\
&=j(t,l)E(e^{ilB_s}-1)
\end{aligned}$$

由 Taylor 展开式 $e^{ix}=1+ix-\frac{x^2}{2}+o(x^2)$，我们有(严格一些还需假定 $E|B_t|^3<\infty$)

$$E(e^{ilB_s}-1)=il\cdot EB_s-\frac{l^2}{2}EB_s{}^2+o(EB_s{}^2)=-\frac{1}{2}l^2s+o(s)\,.$$

对（E．1）式两端除以 s，并令 $s\to 0$ 便得

$$\frac{\partial j}{\partial t}(t,l)=-\frac{1}{2}l^2 j(t,l)\,.$$

这是关于 t 的一个常微分方程，解这个方程可得

$$j(t,l)=j(0,l)e^{-\frac{1}{2}l^2t}\,.$$

再注意到 $j(0,l)=E(e^{ilB_0})=1$，就得到

$$j(t,l)=e^{-\frac{1}{2}l^2t}=\int_{-\infty}^{\infty}e^{ilx}\frac{e^{-\frac{x^2}{2t}}}{\sqrt{2p\cdot t}}dx \qquad \text{(E．2)}$$

从（E．2）可以看出 $j(t,l)=e^{-\frac{1}{2}l^2t}$ 恰是正态分布 $N(0.t)$ 的特征函数．因此 $B_t, B_{t+s}-B_s\sim N(0.t)$（注意特征函数与分布之间的对应是双射）．另外，显见有 $E|B_t-B_s|^2=|t-s|$.

3．2　Brown 运动 (数学模型)

定义３．１７　时齐的独立`增量过程 B_t 称为 Brown 运动，如果它满足

（1）$\forall s, B_{t+s}-B_s\sim N(0,t)$;

（2）对于固定的 w (基本事件)，轨道 $B_t(w)$ 是 t 的连续函数 (称为轨道连续的随机过程).

一般地 Brown 运动也可以对初值 B_0 不加什么限制，也就是说，它可以是任何随机变量.

定义３.１８ 在$B_0 = x$的条件下，B_t的条件分布密度$p_{B_t}(y \mid B_0 = x)$，称为 Brown 运动的转移密度，我们把它改记为$b(t,x,y)$．于是有：在$B_0 = x$的条件下，$B_t - x \sim N(0,t)$．因而

$$b(t,x,y) = \frac{1}{\sqrt{2p \cdot t}} e^{-\frac{(y-x)^2}{2t}}. \qquad (3.\ 10)$$

推论３．１９ Brown 运动对于不同的t对应的转移密度族满足以下的关系

$$b(t+s,x,y) = \int b(t,x,u)b(s,u,y)du \qquad (3.\ 11)$$

这个关系实际上是转移密度的全概率公式，称为 Chapman-Kolmogorov 方程.

因为 Brown 是独立增量过程，所以它有 Markov 性质（无后效性质），即对$0 \le s_m < \cdots < s_1 < s < s+t$，在$B_s = x, B_{s_1} = x_1, \cdots, B_{s_m} = x_m$的条件下，$B_{s+t}$的条件密度

$$p_{B_{s+t}}(y \mid B_s = x, B_{s_1} = x_1, \cdots, B_{s_m} = x_m) = p_{B_t}(y \mid B_0 = x) = b(t,x,y)\text{，} \qquad (3.12)$$

即是 Brown 的转移密度．具有 Markov 性质的随机过程称为 Markov 过程. Brown 运动的转移函数是与时刻s无关的，所以称为时齐的 Markov 过程． Markov 过程相当于动力系统的随机版本，动力系统是在已知现在的条件下，过去对于将来会发生什么是不起影响的．而 Markov 过程是在已知现在的条件下，过去对于将来会发生什么的概率规律不起影响． Markov 过程的时齐性对应于确定性的定常的动力系统.

[警告] "现在的取值" 只能是$B_s = x$，而不能是$B_s \in A$。

Brown 运动的联合分布密度 设B_0的分布密度为$j(x_0)$，那么由独立增量性，对于$0 < t_1 < \cdots < t_n, (B_0, B_{t_1}, \cdots, B_{t_n})$的联合分布函数为

$$P(B_0 \le x_0, B_{t_1} \le x_1, \cdots, B_{t_n} \le x_n)$$

$$= \int_{-\infty}^{x_0}\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n} j(x_0)b(t_1,x_0,x_1)\cdots b(t_n - t_{n-1}, x_{n-1}, x_n)dx_0 \cdots dx_n .$$

所以，对于$0 < t_1 < \cdots < t_n, (B_0, B_{t_1}, \cdots, B_{t_n})$的联合密度是初始密度与各个时刻间的转移密度的乘积，即

$$j(x_0)b(t_1,x_0,x_1)\cdots b(t_n - t_{n-1}, x_{n-1}, x_n). \qquad (3.13)$$

3. 3 Brown 运动的简单性质

本段中将给出 Brown 运动的一些最简单的分布性质与轨道性质，不妨假定$B_0 = 0$．

Brown 运动是一个 Gauss 过程， 因此它的分布特性由其均值向量和协方差矩阵完全确定．由$EB_t = 0$及 Brown 运动的独立增量性可知，当$s \le t$时有

$$Cov(B_s,B_t)=Cov(B_s,B_s+B_t-B_s)=Cov(B_s,B_s)+Cov(B_s,B_t-B_s)=s\text{。}$$

对称地，当 $t>s$ **时，** $Cov(B_s,B_t)=t$ **，从而，我们有**

$$Cov(B_s,B_t)=E(B_sB_t)=s\wedge t\overset{\Delta}{=}\min(s,t)\text{。} \qquad (3。14)$$

容易验证 Brown **运动具有如下的平移不变性和尺度不变性：**

（1）**平移不变性：若**$\{B_t,t\ge 0\}$ **为** Brown **运动，则**$\{B_{t+a}-B_a,t\ge 0\}$（a **为常数）也是** Brown **运动；**

（2）**尺度不变性：若**$\{B_t,t\ge 0\}$ **为** Brown **运动，则**$\{\frac{B_{ct}}{\sqrt{c}},t\ge 0\}$（$c$ **为正常数）也是** Brown **运动.**

此外, Brown **运动还有如下的重要的轨道性质．　由** Brown **运动**$\{B_t,t\ge 0\}$ **的定义，它的轨道是连续的．可是这些轨道却都是处处都不可微的函数． 后一个性质的物理实质在于，粒子在每一瞬间都会受到介质中的分子的碰撞，碰撞后的粒子改变运动方向，因而没有速度**. **从数学的角度看，由于** $B_{t+h}-B_t\sim N(0,h)$**，故而对任意给定的大的正数** M **有**

$$P(|\frac{B_{t+h}-B_t}{h}|\ge M)=2[1-\Phi(M\sqrt{h})]\xrightarrow{h\to 0}1,$$

其中 $\Phi(x)=\frac{1}{\sqrt{2p}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt$ **是**$N(0,1)$ **的分布函数**. **粗略地说，**Brown **运动在任一点**t **处存在有限导数（此时差商**$\frac{B_{t+h}-B_t}{h}$**必然有限）的概率为** 0.

经过严格概率论的进一步分析，还可以知道， Brown **运动在几乎每条轨道上的任意一点处，其导数均不存在．**

3．4　Brown **运动的反射原理及首达性质**

Brown **运动的反射原理**　Brown **运动具有"镜面反射对称性"，称为 Brown 运动的反射原理**. **它可以直观地叙述为**

$$P(Brown\text{运动在}[0,t]\text{到过}x,\text{且}B_t\le x)=P(Brown\text{运动在}[0,t]\text{到过}x,\text{且}B_t\ge x).$$

（3。15）'

设想在 x **处放上一个镜子，则（３．１５）的含义为：**Brown **运动的概率关于此镜子具有对称性**. **这个等式的证明需要较多的知识，在本书中我们不予给出**.. **注意到（３．１５）右方就是** $P(B_t\ge x)$，**如果记**

$$B_t^*\overset{\Delta}{=}\max_{s\le t}B_s,$$

那么（３．１５）可以用准确的数学语言叙述如下

定理 3．２０　(**Brown 的反射原理**)

$$P(B_t^* \ge x, B_t \le x) = P(B_t \ge x). \quad (3。15)$$

Brown **运动的首达性质**　　**在应用中常会遇到** Brown **运动首达** a **的时刻（记为** T_a **），即**

$$T_a \stackrel{\Delta}{=} \min\{t > 0 : B_t = a\}.$$

显见 T_a **是一随机变量. 而反射原理可以在直观上理解为 ：**

当 $t > T_x$ **时，** $B_{T_x} - (B_t - B_{T_x})$ **与** $B_{T_x} + (B_t - B_{T_x})$ **同分布，即** $2B_{T_x} - B_t$ **与** B_t **同分布.**

由此可以设想， 在时刻 T_a **以后对** Brown **运动接上** $2B_{T_a} - B_t$ **，是否也能是一个（是另一个!）Brown 运动呢 ？ 概率论理论证明了这个猜测是正确的. 即对于任意** $t \ge 0$ **，如下定义的随机过程**

$$\check{B}_t \stackrel{\Delta}{=} \begin{cases} B_t & (t < T_a) \\ 2B_{T_a} - B_t & (t \ge T_a) \end{cases} = \begin{cases} B_t & (t < T_a) \\ 2a - B_t & (t \ge T_a) \end{cases}$$

与 B_t **同分布 （指一切有限维分布都相同）. 也就是说** $\check{B}_t$ **也是一个 Brown 运动。 经过更细致的分析， 还可以导出下述定理**

定理 3．21 **随机过程** (T_a, B_t) **与随机过程** $(T_a, \check{B}_t)$ **具有相同的有限维分布族.**

应用反射原理可以得到 T_a **的分布.**

定理 3．22　 **若** $B_0 = 0$**，** **则** Brown **运动** $\{B_t, t \ge 0\}$ **首达** $a(>0)$ **的首达时** T_a **的分布函数的表达式为**

$$P(T_a \le t) = \sqrt{\frac{2}{p}} \int_{\frac{a}{\sqrt{t}}}^{\infty} e^{-\frac{x^2}{2}} dx = 2(1 - \Phi(\frac{a}{\sqrt{t}})),$$

因此，其密度函数为

$$f_{T_a}(t) = \frac{a}{\sqrt{2\pi t^3}} e^{-\frac{a^2}{2t}} I_{[0,\infty)}(t).$$

证明　**注意由** Brown **运动的连续性知** $\{T_a \le t\} = \{B_t^* \ge a\}$. **应用反射原理，我们有**

$$P(T_a \le t) = P(B_t^* \ge a) = P(B_t^* \ge a, B_t \le a) + P(B_t > a)$$

$$= 2P(B_t \ge a) = \frac{2}{\sqrt{2\pi t}} \int_a^{\infty} e^{-\frac{x^2}{2t}} dx = 2(1 - \Phi(\frac{a}{\sqrt{t}})).$$

［**注１**］**当** $a<0$ **时，由** $B_0 = 0$ **及对称性可知，** T_a **的分布与** T_{-a} **的分布相同。因此，对任意实数** a **，有**

$$P(T_a \le t) = \sqrt{\frac{2}{p}} \int_{\frac{|a|}{\sqrt{t}}}^{\infty} e^{-\frac{x^2}{2}} dx = 2(1 - \Phi(\frac{|a|}{\sqrt{t}})).$$

[**注2**] **分布密度为**

$$p(x) = \sqrt{\frac{l}{2px^3}} \exp(-\frac{l}{2m^2} \frac{(x-m)^2}{x}) I_{[0,\infty)}(x)$$

的分布称为逆正态分布(**逆 Gauss 分布，反正态分布**)**. 于是** Brown **的首达时分布相当于参数** $\lambda = a^2, m = \infty$ **的逆正态分布.**

推论3．23 T_a **几乎处处有限，即** $P(T_a < \infty) = 1$

证明　　我们有

$$P(T_a < \infty) = \lim_{t \to \infty} P(T_a \le t) = \sqrt{\frac{2}{p}} \int_0^{\infty} e^{-\frac{x^2}{2}} dx = 1. \qquad ?$$

它的概率意义为：无论 a **多大，**Brown **运动总是以概率** 1 **在有限时间内到达** a **。**

推论　3．24 $ET_a = +\infty$

证明　利用 T_a **的密度函数得到**

$$ET_a = \int_0^{\infty} t f_{T_a}(t) dt = \int_0^{\infty} \frac{ta}{\sqrt{2pt^3}} e^{-\frac{a^2}{2t}} dt = \int_0^{\infty} \frac{a}{\sqrt{2pt}} e^{-\frac{a^2}{2t}} dt = +\infty. \qquad ?$$

此推论表明, 虽然 Brown **运动几乎所有的轨道首次到达** a **的时间有限,但是对于所有的轨道到达** a **点的平均时间却是无穷大.**

有关的另一个使人感兴趣的随机变量是 Brown **运动在** $[0,t]$ **中达到的最大值** $B_t^* = \max_{0 \le s \le t} B_s$**，当** $a \ge 0$ **时我们有**

$$P(\max_{0 \le s \le t} B_s \ge a) = P(T_a \le t) = \sqrt{\frac{2}{p}} \int_{\frac{a}{\sqrt{t}}}^{\infty} e^{-\frac{x^2}{2}} dx = 2(1 - \Phi(\frac{a}{\sqrt{t}})).$$

因此得到下面的推论

推论3．25 $\max_{0 \le s \le t} B_s$ **的密度函数为**

$$f_{\max}(a) = \frac{2}{\sqrt{2pt}} e^{-\frac{a^2}{2t}}, \quad a \ge 0.$$

定理3．26 (Brown **运动的的反正弦律**) **设** B_t **为** Brown **运动**, $B_0 = 0$, $0 < s < t$, **则**

$$P(B_u \neq 0 : \forall u \in (s,t)) = \frac{2}{p} \arcsin \sqrt{\frac{s}{t}} .$$

证明纲要 将待求概率的事件的对立事件记为 A，即 $A = \{\exists u \in (s,t) \text{使} B_u = 0\}$． 由推广的全概率公式推出

$$P(A) = \int P(A \mid B_s = a) \frac{1}{\sqrt{2ps}} e^{-\frac{s^2}{2}} ds$$

对于 $a > 0$，利用 Brown 运动的时齐性，空间平移不变性(参见习题 3 的 23 题)和对称性，对于 Brown 运动首达 0 的 t_0，由 Brown 运动的性质，我们有

$$P(A \mid B_s = -a) = P(A \mid B_s = a) = P(\min{}_{s<u<t} B_u \le 0 \mid B_s = a)$$

$$= P(\min{}_{0<u<t-s} B_u \le 0 \mid B_0 = a) = P(\max{}_{0<u<t-s} B_u \ge a \mid B_0 = 0)$$

$$= P(t_a \le t - s \mid B_0 = a) = \frac{a}{\sqrt{2p}} \int_0^{t-s} v^{-\frac{3}{2}} e^{-\frac{s^2}{2v}} dv .$$

于是

$$P(A) = \int \frac{a}{\sqrt{2p}} \int_0^{t-s} v^{-\frac{3}{2}} e^{-\frac{s^2}{2v}} dv \frac{1}{\sqrt{2ps}} e^{-\frac{s^2}{2}} ds$$

$$= \cdots \text{ 经过化简后} \cdots = \frac{1}{p\sqrt{s}} \int_0^{t-s} \frac{du}{\sqrt{u}(u+s)} = \qquad (\text{令 } u = sv^2)$$

$$= \cdots = \frac{2}{p} arctan \sqrt{\frac{t-s}{s}} = \frac{2}{p} \arccos \sqrt{\frac{s}{t}} .$$

(因为如果记 $J = arctan\sqrt{\frac{t-s}{s}}$，那么 $\frac{t-s}{s} = tan^2 J$．于是 $\cos J = \frac{1}{\sqrt{1+\frac{t-s}{s}}} = \sqrt{\frac{s}{t}}$).

所以

$$P(B_u \neq 0 : \forall u \in (s,t)) = 1 - P(A) = 1 - \frac{2}{p} \arccos \sqrt{\frac{s}{t}} = \frac{2}{p} \arcsin \sqrt{\frac{s}{t}} . \qquad ?$$

Levy 曾在 20 世纪 40 年代，对 Brown 运动作了奠基性的研究．此后 Brown 运动的研究始终十分活跃，方兴未艾，至今还不断地推陈出新．Brown 运动还有许多非常有趣的性质，在本书中不宜涉及更多．

3．5 与 Brown 运动有关的几个简单随机过程

吸附 Brown 运动与吸收 Brown 运动

设 $a > 0, T_a$ 为 Brown 运动 B_t 首达 a 的时刻．令 $B_t^{(a)} = \begin{cases} B_t & (t < T_a) \\ a & (t \ge T_a) \end{cases}$，称为在 a 吸附的吸附

Brown 运动。它的分布函数是既有离散部分又有连续部分的混合分布. 下面我们来推导这个分布函数. 用 (T_a, B_t) 与 $(T_a, \check{B}_t)$ 有相同的联合分布，在 $x < a$ 时得到

$$P(B_t^{(a)} \le x) = P(B_t \le x, B_t^* < a) = P(B_t \le x) - P(B_t \le x, B_t^* \ge a)$$

$$= P(B_t \le x) - P(B_t \le x, T_a \le t) = P(B_t \le x) - P(\check{B}_t \le x, T_a \le t)$$

$$= P(B_t \le x) - P(2a - B_t \le x, T_a \le t).$$

再注意到 $\{2a - B_t \le x\} \subset \{T_a \le t\}$， 因此在 $x < a$ 时有

$$P(B_t^{(a)} \le x) = P(B_t \le x) - P(2a - B_t \le x) = P(B_t \le x) - P(B_t \ge 2a - x)$$

$$= P(B_t \le x) - P(B_t \le x - 2a) = P(x - 2a < B_t \le x)。$$

另一方面我们还有 $P(B_t^{(a)} = a) = P(T_a \le t) = 2P(B_t \ge a)$. 两者统一起来就有

定理 3．27

$$P(B_t^{(a)} \le x) = \begin{cases} P(x - 2a < B_t \le x) & (x < a) \\ 1 & (x \ge a) \end{cases}.$$

即吸附 Brown 运动 $B_t^{(a)}$ 的分布函数在 $x = a$ 处不连续， 它有一个跳跃量 $2P(B_t \ge a)$.

［注1］ Brown 运动在首达 a 以前的部分 $\{B_t : t < T_a\}$, 称为在 a 点吸收的吸收 Brown 运动．确切地说， 吸收 Brown 运动不是一个"完整的"随机过程，它在 $t \ge T_a$ 后就消失了． 把吸收 Brown 运动与吸附 Brown 运动 $B_t^{(a)}$ 进行比较, 就得到在 a 点吸收的吸收 Brown 运动取值的概率规律为

$$P(B_t \le x, t < T_a) = P(x - 2a < B_t \le x), (x < a).$$

［注2］ 设 $x_0 > 0, B_0 = 0$． 通常把 $x_t \stackrel{\Delta}{=} x_0 + B_t$ 称为 "从 x_0 出发的 Brown 运动"．易见它的转移密度为 $p(t, x, y) = b(t, x - x_0, y - x_0)$，且 ξ_t 首达 0 的时刻正是 B_t 首达 $-x_0$ 的时刻. 记 x_t 首达 0 点的时刻为 T_0， 那么 x_t 在到达 0 点以前的部分, 称为在 0 点吸收的吸收 Brown 运动.

0 点反射的反射 Brown 运动

$B_t^{(r)} \stackrel{\Delta}{=} |B_t|$ 称为（在 0 点反射的）反射 Brown 运动。我们有

$$E|B_t| = \sqrt{\frac{2t}{p}},\ Var|B_t| = (1 - \frac{2}{p})t.$$

积分 Brown 运动

$\xi_t \stackrel{\Delta}{=} \int_0^t B_s ds$ 称为积分 Brown 运动，它不再具有 Markov 性．但是不难证明，二元随机过程$(\int_0^t B_s ds, B_t)$却是 Markov 过程．

3. 6　漂移 Brown 运动

定义３．２８　$x_t = B_t + m \cdot t$ 称为具有常数漂移m的漂移 Brown 运动．这是 Brown 运动加上一个线性趋势项．于是（以下诸叙述请读者补出证明）

(1)　漂移 Brown 的转移密度，即条件分布密度 $p_{x_{s+t}}(y \mid x_s = x)$，是 $p(t,x,y) = b(t,x,y-mt)$．

(2)　漂移 Brown 运动是时齐的 Markov 过程，即它满足

$$p_{x_{s+t}}(y \mid x_s = x, x_{s_1} = x_1, \cdots, x_{s_m} = x_m) = p_{x_t}(y \mid x_0 = x).$$

(3)．对于$0 < t_1 < \cdots < t_n$而言，随机变量组$(x_0, x_{t_1}, \cdots, x_{t_n})$的联合密度是初始密度$j(x_0)$与各个时刻间的转移密度的乘积，即

$$j(x_0) b(t_1, x_0, x_1 - mt_1) \cdots b(t_n - t_{n-1}, x_{n-1}, x_n - m(t_n - t_{n-1})). \qquad (3.14)$$

漂移 Brown 运动可以作如下的随机模拟：

设 $\{X_k\}$独立同分布，$X_k \sim \begin{pmatrix} -1 & 1 \\ 1-p & p \end{pmatrix}$。定义 $X_t^{\Delta} = \Delta x(X_1 + \cdots + X_{[\frac{t}{\Delta t}]})$，则

$$EX_t^{\Delta} = \Delta x[\frac{t}{\Delta t}](2p-1)，Var(X_t^{\Delta}) = (\Delta x)^2[\frac{t}{\Delta t}][1-(2p-1)^2]。$$

对于充分小的数$h > 0$，如果取 $\Delta t = h, \Delta x = \sqrt{h}$，并且对于给定的常数$m$，选取$p$使 $2p - 1 = m\sqrt{h}$（即 $p = \frac{1}{2}(m\sqrt{h} + 1)$），那么当 $h \to 0$ 时，$EX_t^{\Delta} \to m \cdot t$，$Var(X_t^{\Delta}) \to t$．由中心极限定理可得 X_t^{Δ}的近似分布为 $N(m \cdot t, t)$．进一步用多维的中心极限定理，可以证明X_t^{Δ}的所有的有限维分布，都收敛于漂移 Brown 运动的相应的有限维分布（在较为深入的随机过程理论课程中，还可以证明：在$h \to 0$时，X_t^{Δ}作为随机过程整体在某种意义下的极限为漂移 Brown 运动）．

特别地，取$m = 0$，$p = \frac{1}{2}$，则相应的极限就是 Brown 运动．

3. 7　几何 Brown 运动

定义３．２９　对数为漂移 Brown 运动的随机过程，称为几何 Brown 运动．即$x_t = e^{B_t + mt}$称为几何 Brown 运动．有些文献上也称$x_t = e^{sB_t + mt}, (s > 0)$为几何 Brown 运动．

几何Brown运动不是独立增量过程. 但是, 由于指数函数是严格增函数, 所以它仍是Markov过程. (易见, 若x_t是 Markov 过程, f 为严格单调函数, 则$V_t \overset{\Delta}{=} f(x_t)$ 也是 Markov 过程).

在金融证券中, 最简单的变量如股票价格, 外汇汇率, 股票指数价格, 期货价格等等, 都是随机过程, 在金融中统称为标的变量(underlying variable). 大量实证表明, 用几何Brown 运动作为这种标的随机过程的粗近似, 能提供明显的参考作用. 在金融中的几何Brown 运动常显现为

$$S_t = e^{\sigma B_t + (\mu - \frac{\sigma^2}{2})t}, (\sigma > 0).$$

以此作为数学模型，称为 Black-Scholes 模型. 参数m称为此标的变量的收益率。我们将证明, 在无套利的市场中确定与标的变量有关的金融产品的合理定价时, m 应该取为银行的无风险利率r. 这样的模型$S_t = e^{\sigma B_t + (r - \frac{\sigma^2}{2})t}, (\sigma > 0)$称为风险中性的 Black-Scholes 模型. 这时参数s 就成为唯一的能起作用的模型参数, 称为波动率(volatility), 它代表市场的随机波动. 在应用中就需要利用实际数据资料来估计波动率s.

假定标的变量在[0,T]区间按步长Δt的采样为: $S_{n\Delta t}(0 \le n \le N, T = N\Delta t)$. 那么,

$\ln \frac{S_{n\Delta t}}{S_{(n-1)\Delta t}}, (N \ge n > 0)$相互独立，而且与$\sigma B_{\Delta t} + (r - \frac{\sigma^2}{2})\Delta t$具有相同的分布, 即其分布为

$N((r - \frac{s^2}{2})\Delta t, s^2 \Delta t)$. 于是$s^2$的估计可取为

$$\hat{s}^2 = \frac{1}{\Delta t(N-1)} \sum_{n=1}^{N} (\ln \frac{S_{n\Delta t}}{S_{(n-1)\Delta t}} - \frac{1}{N} \ln \frac{S_T}{S_0})^2. \qquad (3.\ 15)$$

4. 简单随机徘徊

通常的教科书中， 常把简单随机徘徊常放在 Brown 运动之前, 用简单随机徘徊作为Brown 运动的近似. 然而在本书中, 我们考虑到连续情形比离散情形更易于处理, 所以把简单随机徘徊放在 Brown 运动后面.

4．1．双侧吸收壁的吸收概率

假定两人参加博弈．开始时甲有资本a元，乙有资本b元，每次１元，直至其中一人输光就停止，不容许借贷．求甲输光的概率p_a和乙输光的概率q_b．

此问题可用两端带有吸收壁的随机徘徊来描述． 假定质点在每隔一个单位时间以概率p和q分别向右和向左移动一个单位距离，且每次移动之间是相互独立的．再假定质点在开始时刻t=0时位于$n(0 \le n \le a+b)$，且在0及$a+b$处各设一个吸收壁，即质点一旦到达0或$a+b$，就永远停在该处．我们来求质点最终在0被吸收的概率p_n．由定义显然有$p_0 = 1, p_{a+b} = 0$．我们分析在n处的质点要被0吸收，总共只有两种可能：即先向右移动一格而最后被0吸收，或先向左移动一格而最后被0吸收．用全概率公式就得到差分方程

$$p_n = pp_{n+1} + qp_{n-1}.$$

它等价于 $p(p_{n+1} - p_n) = q(p_n - p_{n-1})$**，即**

$$p_{n+1} - p_n = \frac{q}{p}(p_n - p_{n-1}).$$

下面分两种情形分别考虑．

（1） $p = q$ **情形．这时** $p_{n+1} - p_n =$**常数，记为** c**．于是** $p_n = p_0 + nc = 1 + nc$**．再由** $0 = p_{a+b} = 1 + (a+b)c$ **得到** $c = -\frac{1}{a+b}$**．故而** $p_n = 1 - \frac{n}{a+b}$**．所以** a **出发而在**0**被吸收的概率为** $p_a = 1 - \frac{a}{a+b} = \frac{b}{a+b}$**．同理可得** $q_b = \frac{a}{a+b}$**．**

（2） $p \neq q$ **情形．由**$p_0=1$**得到**

$$p_{n+1} - p_n = (\frac{q}{p})^n (p_1 - 1),$$

于是

$$p_{a+b} - p_n = \sum_{k=n}^{a+b-1}(p_{k+1} - p_k) = \sum_{k=n}^{a+b-1}(\frac{q}{p})^n (p_1 - 1) = (p_1 - 1)\frac{(\frac{q}{p})^n - (\frac{q}{p})^{a+b}}{1 - \frac{q}{p}}.$$

再由 $p_{a+b}=0$**推出**

$$p_n = (1 - p_1)\frac{(\frac{q}{p})^n - (\frac{q}{p})^{a+b}}{1 - \frac{q}{p}},$$

又由于$p_0=1$**，便得**

$$1 = (1 - p_1)\frac{1 - (\frac{q}{p})^{a+b}}{1 - \frac{q}{p}},$$

最后得到

$$p_n = \frac{(\frac{q}{p})^n - (\frac{q}{p})^{a+b}}{1 - (\frac{q}{p})^{a+b}}。$$

因此

$$p_a = \frac{(\frac{q}{p})^a - (\frac{q}{p})^{a+b}}{1 - (\frac{q}{p})^{a+b}} = \frac{1 - (\frac{p}{q})^b}{1 - (\frac{p}{q})^{a+b}}.$$

用同样的方法可得

$$q_b = 1 - p_a = \frac{1-(\frac{q}{p})^a}{1-(\frac{q}{p})^{a+b}}.$$

4．2　随机徘徊的对称原理

定理 3．3 0 (**随机徘徊的对称原理**)　**设** S_n **为随机徘徊**，　$S_n = X_1 + \cdots + X_n$，X_k **独立同分布**．**那么我们有（注意：正反方向的起始分别为** 0 **与** S_n **）**

$$P(S_i > 0, (i<n), S_n \in [a,b]) = P(S_i < S_n, (i<n), S_n \in [a,b]). \qquad (3.16)$$

证明　记 $X_1, \cdots, X_n$ **的逆向排列为** $\hat{X}_1, \cdots, \hat{X}_n$，**即** $\hat{X}_k \stackrel{\Delta}{=} X_{n-k+1}$　$(k \le n)$．**于是** $\hat{S}_n \stackrel{\Delta}{=} \hat{X}_1 + \cdots + \hat{X}_n$ **也是简单随机徘徊，而且有**

$$\hat{S}_i = \hat{X}_1 + \cdots + \hat{X}_i = X_n + \cdots + X_{n-i+1} = S_n - S_{n-i},\ \hat{S}_n = S_n.$$

由此就得到

$$P(S_i > 0, i<n, S_n \in [a,b]) = P(\hat{S}_i > 0, i<n, \hat{S}_n \in [a,b]) = P(S_n > S_i, i<n, S_n \in [a,b]).$$

4．2　**随机徘徊的首达时刻**

引理 3．3 1　在满足 $x_1 + \cdots + x_n = m, x_i < 2, (i \le n)$ **的整数解**(**可以取负值**) $x_1, \cdots, x_n$ **的** n **个循环排列中，恰有** m **个循环排列满足：部分和的前** $n-1$ **个都小于** m．(**即：满足条件**

$$“x_{i_1} + \cdots + x_{i_p} < m, \quad (p<n)”$$

的循环排列：$x_{i_1}, \cdots, x_{i_n}$ **只有** m **个**).

证明　首先注意到，因为总和为 m，**所以满足条件要求的解** $(x_1, \cdots, x_n)$ **必须有** $x_n = 1$．**再则，** $x_1, \cdots, x_n$ **的部分和共** n **个，把它们的最大值记为** M．**假定其中第** l **个部分和是达到** M **的那些部分和中的第一个．那么也必定有** $x_l = 1$ (**否则有** $x_l < 1$，**于是前面的部分和中就必有达到** M **者，这就出现了矛盾**)．**从而由** $x_1 + \cdots + x_n = m, x_i < 2, (i \le n)$ **的解得到的** $(x_{l+1}, x_{l+2}, \cdots x_n, x_1, \cdots, x_l)$ **就是满足上述要求的一个循环排列**.

往证这样的循环排列一共只有 m **个．为了叙述简单，下面不妨假定这个循环排列就是** $(x_1, \cdots, x_n)$ **自己．对于** $j = 1, \cdots, m$，**记** $p_j = \min\{k : x_1 + \cdots + x_k = j\}$，**它是部分和首达** j **的时刻．由假定可知**

$p_m = n$. 可见和数 $x_1 + \cdots + x_n$ 达到 m 是由分别达到累计和为 1 的 m 段 $\{x_{p_i+1}, x_{p_i+2}, \cdots, x_{p_{i+1}}\}$ 逐步实现的 ($\sum_{j=p_i+1}^{s_i} x_j < 1, (\forall s_i < p_{i+1}, p_0 = 0, i = 0,1,\cdots, n-1)$). 因此, 除 $(x_1,\cdots,x_n)$ 自己外, 只有以 x_{p_j+1} 开始的排列且当 $1 \le j \le m-1$ 时, 才满足我们的要求. 引理证毕.

定理 3．3 2 (Dwass – Dinges 定理) 设 S_n 为从 x 出发的随机徘徊

$S_n = x + X_1 + \cdots + X_n$, 其中 X_i 独立同分布, 且取值于 $\{1,0,-1,-2,\cdots\}$. $S_0 = x$ (整数). 那么, 在 $m > x$ 时有

$$P(S_i < m,(i < n), S_n = m) = \frac{m-x}{n} P(S_n = m). \qquad (3.17)$$

证明 不妨假定 $x = 0$ (否则可用 $m - x$ 代替 m). 把 $(X_1,\cdots,X_n)$ 的 n 个循环排列的联合 $\{(X_1,\cdots,X_n),(X_2,\cdots,X_n,X_1),\cdots,(X_n,X_1,\cdots,X_{n-1})\}$ 所出现的一切可能放在一起, 作为样本空间. 由独立同分布性可知其中 n 部分的任意一部分内所有随机变量的和都是具有相等概率的, 我们可以把它们看成为 "广义的" 基本事件. 于是我们可以用推广的古典计算方法来计算条件概率 $P(S_i < m,(i < n) \mid S_n = m)$. 再由引理３．３１可知, 在上面的 "广义的" 基本事件中, 有利于我们要算的(条件)概率的恰有 m 个, 因此, 这个(条件)概率应为 $\frac{m}{n}$. 这就是我们要证明的结论.

4. 3 简单随机徘徊与首达时

若 S_n 为简单随机徘徊 $S_n = X_1 + \cdots + X_n$, 其中 $X_i \sim \begin{pmatrix} -1 & 1 \\ q & p \end{pmatrix}$, 且相互独立。那么利用 Dwass-Dinges 定理得到, 对于 S_n 首达 1 的时刻 T_1 有

$$P(T_1 = 2n+1) = \frac{1}{2n+1} P(S_{2n+1} = 1) = \frac{1}{n+1} C_{2n}^n p^{n+1} q^n. \qquad (3.18)$$

于是 T_1 的矩母函数 (注意此时可能有 $P(T_1 < \infty) < 1$, 所以此时矩母函数在 1 处的值就会小于 1) 为 $pA_1(pqz)$, 其中 (T_1 只可能取奇数)

$$A_1(z) = \sum_{n=0}^{\infty} \frac{1}{n+1} C_{2n}^n z^n = \sum_{n=0}^{\infty} \frac{(2n)!}{n!(n+1)!} z^n.$$

注意到 $1 - 2zA_1(z)$ 的展开式中 z^n 的系数为

$$-2\frac{(2(n-1))!}{(n-1)!n!}=-2\cdot 2^{n-1}\frac{(2n-3)(2n-5)\cdots 3\cdot 1}{n!}=(-4)^n\frac{\frac{1}{2}(\frac{1}{2}-1)(\frac{1}{2}-2)\cdots(\frac{1}{2}-(n-1))}{n!}$$

把它与广义 Newton 展开式

$$(1+y)^a=\sum_{n=0}^{\infty}\frac{a(a-1)\cdots(a-(n-1))}{n!}y^n$$

对照，便得到

$$1-2zA_1(z)=(1-4z)^{\frac{1}{2}}.$$

即 $A_1(z)$ 有显式表示

$$A_1(z)=\frac{1-\sqrt{1-4z}}{2z}. \qquad (3.18)$$

从而得到

$$P(T_1<\infty)=pA_1(pq)=\frac{1-|p-q|}{2q}=\begin{cases}1 & (p\ge q)\\ \frac{p}{q} & (p<q)\end{cases}, \qquad (3.19)$$

$$ET_1=pA_1'(pq)=\begin{cases}\frac{1}{p-q} & (p>q)\\ \infty & (p\le q)\end{cases}. \qquad (3.20)$$

进一步记首达 m 的时刻为 T_m，同样用 Dwass-Dinges 定理可得到

$$P(T_m=2n+m)=\frac{m}{2n+m}P(S_{2n+m}=m)=\frac{1}{n+m}C_{2n+m-1}^{n}p^{n+m}q^n. \qquad (3.21)$$

再注意简单随机徘徊 S_n 从 k 到 $k+1$ 的转移规律与从 0 到 1 是一样的，由 Markov 性便得到

下面的定理

定理３．３３

$$P(T_m<\infty)=\begin{cases}1 & (p\ge q)\\ (\frac{p}{q})^m & (p<q)\end{cases}, \qquad (3.22)$$

$$ET_m=\begin{cases}\frac{m}{p-q} & (p>q)\\ \infty & (p\le q)\end{cases}. \qquad (3.23)$$

习题 3

1. 设 $\{N(t),t\ge 0\}$ 服从强度为 l 的 Poisson 过程，$s<t$．求 $E(N_s\mid N_t),E(N_t\mid N_s)$ 以及协方差

函数$Cov(N(t), N(t+s))$.

2. 设 N_t 是 Poisson 过程. $0 \le s \le t$, 证明 $P\{N_s \le N_t\}=1$, 且 $\lim_{t\to s} P\{N_t - N_s > e\} = 0$.

3. 设 N_t 为 Poisson 过程, $s_1 < \cdots < s_m < t$. 证明在 $N_t = n$ 的条件下, $(N_{s_1}, \cdots, N_{s_m})$ 的条件联合分布为多项分布.

4. 假定海浪引起的多次冲击形成一个参数为 l 的指数流 $\{t_k\}$, 第 k 次冲击对某个设备造成的瞬时随机损失为 D_k. 设. $\{D_k, k \ge 1\}$ 为独立同分布, 且与 $\{t_k\}$ 独立. 于是海浪在 t 时刻对设备造成的总损失为

$x_t = \sum_{k=1}^{N_t} D_k e^{-a(t-t_k)}$, 其中 N_t 是 $\{t_k\}$ 的计数过程. 求 $Ex_t, Var(x_t)$.

5. 设正随机变量 Λ 具有密度 $g(l)$. 已知在 $\Lambda = l$ 的条件下, 随机过程 N_t 的条件分布是强度为 l 的 Poisson 分布. 求 $(N_{t_1}, \cdots, N_{t_n})$ 的(联合)分布. 证明 N_t 不是独立增量过程, 而是平稳增量过程, 即 $X_t \overset{\Delta}{=} N_{s+t} - N_s$ 的分布与 s 无关, 而且 $(X_{t_1+t}, \cdots, X_{t_n+t})$ 的分布不依赖 t.

6. 设某商场在下午 1 点到 4 点到达的顾客数 X 服从平均速率为每分钟 60 人的 Poisson 过程, 而到商场的人所消费的钱 $h_i \sim B(30,25)$. 假设 h_i 相互独立且与 X 独立. 求商场在该时间段内的平均营业额与营业额的误差.

7. 设 $X_i, (i \le m)$ 们相互独立, 服从参数为 $l_i, (i \le m)$ 的 Poisson 分布, 求在 $X_1 + \cdots + X_m = n$ 的条件下, $(X_1, \cdots, X_m)$ 的条件分布.

8. 设三个机器人甲, 乙, 丙在时刻 $t=0$ 开始工作, 它们正常工作的时间是相互独立的随机变量. 设时刻 t 待修理的机器人在 $(t, t+h)$ 修好能正常工作的概率都是 $l \cdot h + o(h)$.

(1) 求在时刻其中至少一个待修理的概率. (2) 求正好按甲, 乙 丙的次序并且三个全坏的概率.

9. 设 $\{t_n\}$ 为指数流, 其计数过程为 Poisson 过程 N_t. 求在 $N_t = 1$ 的条件下, (t_1, t_2) 的条件密度.

10. 一个从底层上升的电梯, 设 N_i 为在第 i 层进入电梯的人数, 它们是相互独立的, 且 $N_i \sim Poisson_{l_i}$. 由第 i 层进入电梯的每个人独立地以概率 p_{ij} 在第 j 层离开电梯($\sum_{j>i} p_{ij} = 1$). 记 X_j 为在第 j 层离开电梯的人数. 求 EX_j 及 (X_j, X_k) 的分布. (注: 令 N_{ij} 为从第 i 层进入电梯, 并在第 j 层离开电梯的人数. 由分流定理它们相互独立, 而且有 $X_j = N_{1j} + \cdots + N_{j-1,j}$).

11. 设 $N_t^{(i)}\ (i=1,2)$ 是相互独立强度为 l_i 的 Poisson 过程, 分别对应于指数流 $\{t_n^{(i)}\}$. 求

(1) $t^{(1)}_{N_t^{(2)}}$ 及 $N^{(2)}_{t_1}$ 的分布; (2) $E\{t_1^{(1)} \mid N_t^{(1)}+N_t^{(2)})$.

12. 设 $N_t^{(i)}\ (i=1,2)$ 是相互独立强度为 l_i 的 Poisson 过程. 求证 $N_t^{(1)}-N_t^{(2)}$ 不是 Poisson 过程, 而是复合 Poisson 过程， 并求它的特征函数.

13. 若 $x_t = x_0(-1)^{N_t}$, 其中 $\{N_t\}$ 为与 x_0 独立的 Poisson 过程, 而 $x_0 \sim \begin{pmatrix} -1 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$.

（1） 求 x_t 的协方差函数; （2）求 (x_t, x_s) 的特征函数;

14．设随机变量 $t \sim \exp_m$，且与强度为 l 的 Poisson 过程 $\{N_t : t \ge 0\}$ 独立．证明 N_t 服从参数为

$\dfrac{m}{l+m}$ 的几何分布．

15． 设 $\{t_n^{(i)} : n \ge 1\}(i=1,2)$ 是两个参数分别为 l_i 的相互独立的指数流，其对应的计数过程分别为

Poisson 过程 $N_t^{(i)}$．证明 $N^{(2)}_{t^{(1)}_{n+1}} - N^{(2)}_{t^{(1)}_n}$ 服从参数为 $\dfrac{l_2}{l_1+l_2}$ 的几何分布．

16. 某设备由 A,B 两部件构成，可能发生三类故障, 在 $(0,t]$ 中发生第 i 类（$i=1,2,3$）故障的次数为参数是 l_i 的 *Poisson* 过程, 且相互独立. 在第一类故障时 A 不正常. 在第二类故障时 B 不正常. 而在第三类故障时 A,B 都不正常. 设 A,B 的寿命分别为 x,h. 求证 x,h 均为指数分布, (x,h) 服从二维指数分布: $P(x>t,h>s)=\exp\{-l_1 t - l_2 s - l_3(t \vee s)\}$.

17. 设 $N_t^{(k)}$ 为一列相互独立的强度为 l_k 的 Poisson 过程, $l=\sum_{k=1}^{\infty} l_k < \infty$. 求证 $N_t = \sum_{k=1}^{\infty} k N_t^{(k)}$ 是复合 Poisson 过程.

18. 若 N_t 为均值为 $L(t)$ 的非时齐 Poisson 过程, $L^{-1}(t)$ 为 $L(t)$ 的反函数. 问 $N_{L^{-1}(t)}$ 是什么过程 ?

19. 若 N_t 为均值为 $L(t)$ 的非时齐 Poisson 过程. 设 $X_1,\cdots,X_n$ 为 n 个取值在 $[0,t]$ 上, 且是分布函数为 $F(s)=\dfrac{L(s)}{L(t)}$ 的独立同分布随机变量. 证明在 $N_t=n$ 的条件下, 该非时齐 Poisson 过程的各个事故发生的时刻与 $X_{(1)},\cdots,X_{(n)}$ 同分布.

*20 $[0,t]$ 中进入某医院住院部的病人数为均值为 $L(t)$ 的非时齐 Poisson 过程. 设病人的住院时间独立同分布, 具有分布密度 $f(t)$, 且与此非时齐 Poisson 过程独立. 记 $h(t)$ 为时刻 t 住院的人数. 求它的期

望与方差.

21. 设 B_t 为 Brown 运动，分别求 (1)$x_t = sB_t + m\cdot t$, (2)$h_t=(1-t)B_{\frac{t}{1-t}},(t<1)$, (3)$V_t = e^{-bt}B_{e^{2bt}}$

的协方差函数.

22. 设 $s<u<t$, B_t 为 Brown 运动. 求

$$E(B_s\mid B_t),E(B_t\mid B_s)\quad E(B_t\mid B_s,B_u),E(B_s\mid B_u,B_t),E(B_u\mid B_s,B_t).$$

23. 证明 Brown 运动有空间平移不变性:

$$P(B_{t+s}\in A\mid B_s=x)=P(B_{t+s}\in A+z\mid B_s=x+z),$$

其中 $A+z=\{y+z: y\in A\}$.

24. 设 $a<0$， 对于 Brown 运动 $B_t(B_0=0)$，给出在 a 点的反射原理及吸附的 Brown 运动的概率规律.

25. 对于 $0<t<1$，证明在条件 $B_0=B_1=0$ 条件下，B_t 的条件分布密度为 $N(0,t(1-t))$. 再对

$B_0=a, B_1=b$ 的情形推广这个结果.

26. 令 $X_t = tB_{\frac{1}{t}}I_{[0,\infty)}(t)$，$0<t_1<t_2<t_3$，问 X_{t_1} 与 $X_{t_3}-X_{t_2}$ 是否独立？

27. $\forall 0\le t_1<t_2<\cdots<t_{n+1}$. 求条件分布 $P\{B(t_{n+1})\le x\mid B(t_1)=x_1,\cdots,B(t_n)=x_n\}$ 和

$P\{B(t_1)\le x\mid B(t_2)=x_2,\cdots,B(t_{n+1})=x_{n+1}\}$ 的密度，并由此再求相应的条件期望和条件方差.

28. 求 $x_t = B_t - tB_1$ 及 $h_t=(t+1)x_{\frac{t}{t+1}}$ 的分布. 再对 $0<s<t<1$ 求 x_s 与 x_t 的协方差函数与

相关系数

29. 分别求 $|B_t|$，B_t^*，$B_{*t}=\min_{s\le t}B_s$ 及 $D_t=B_t^*-B_t$ 的密度． 求证

$$P\{B_t^*>x\mid D_t=0\}=\exp\left(-x^2/2t\right).$$

30. 记 $S_n=\sum_{k=1}^{2^n}(B_{\frac{k+1}{2^n}}-B_{\frac{k}{2^n}})^2$. 求 ES_2, $E(S_3\mid S_2)$, $E(S_2\mid S_3)$, ES_n, 并证明

$$E(S_{n+1}\mid S_n)=\tfrac{1}{2}(S_n+1),\ E(S_n\mid S_{n+1})=S_{n+1}.$$

31. 对于强度为的 l Poisson 过程 N_t， 证明 $\dfrac{N_t}{t}\xrightarrow{p} l\,(n\to\infty)$.

32．设有 n 个寿命分布均为 Exp_l 的元件同时使用，元件坏了既不修理，也不更新。记第 k 个故障发生的

时刻为 t_k $(k\le n)$，$S_k \overset{def}{=} t_k - t_{k-1}$ $\quad (t_0=0)$.

（1）　求$(t_1,\cdots,t_n)$的联合密度；（2）证明$S_1,\cdots,S_n$独立，且$S_k \sim Exp_{(n-k+1)l}$。

33．对于强度函数为$l(t)$的非时齐的 Poisson 过程N_t，令$t_k = inf\{t : N_t = k\}$，

$S_k \overset{def}{=} t_k - t_{k-1} \quad (t_0 = 0)$。

（1）　求$(t_1,\cdots,t_n)$的联合密度；（2）求$(S_1,\cdots,S_n)$的联合密度。

龚光鲁，钱敏平著 应用随机过程教程及其在算法与智能计算中的应用

清华大学出版社,2003

第 4 章 更新现象及其理论

1. Stieltjes 积分简述

假定 $G(x)$ **为右连续的递增函数．那么，我们可以如微积分中定积分那样定义关于** $G(x)$ **的 Stieltjes 积分**:

$$\int_a^b h(x)dG(x) = \lim_{\substack{n\to\infty \\ \max_i (t_{i+1}^{(n)} - t_i^{(n)})\to 0}} \sum_i h(t_i^{(n)})[G(t_{i+1}^{(n)}) - (G(t_i^{(n)})]\text{，}$$

这里 $\{t_i^{(n)}\}$ **是半开区间** $(a,b]$ **的一个划分．这种积分的运算规律及近似计算，与普通积分基本上类似．最大的不同之处是：由于** $G(x)$ **在某些点上的跃度可以大于零，在这些点上的积分就可能不是** 0. **因此我们通常考虑在半开区间** $(a,b]$ **上的积分．类似地还可以定义瑕积分**

$$\int_{-\infty}^{\infty} h(x)dG(x) = \lim_{\substack{a\to-\infty \\ b\to\infty}} \int_a^b h(x)dG(x)\text{，简记为 } \int h(x)dF(x) \text{ 或 } \int hdF\,.$$

一般地说，Stieltjes **积分更多地用作理论分析，因为除了几种特殊情形或者是它们的混合情形，能计算出积分以外， 通常的** Stieltjes **积分只能用积分和来近似估算．几种能直接计算的特殊情形包括**:

(1) **若** $G(x) = \int_c^x g(u)du$ **，则** $\int_a^b h(x)dG(x) = \int_a^b h(x)g(x)dx$ **化简为普通的积分**.

(2) **若** $G(x)$ **是右连续的递增的阶梯函数，即存在** $\{x_n\} \subset (a,b]$ **，使**

$G(x) = \sum_{x_n \le x} [G(x_n) - G(x_n-)]$ **，则由积分的定义可算出**

$$\int_a^b h(x)dG(x) = \sum_n h(x_n)[G(x_n) - G(x_n-)],$$

这时, Stieltjes **积分就化为无穷级数**.

(3) **对于一个点的集合** $\{x_0\}$ **，有** $\int_{\{x_0\}} h(x)dG(x) = h(x_0)[G(x_0) - G(x_0-)]$.

可见，Stieltjes **积分是一种包容普通的积分与无穷级数，而比之更为广泛的数学模型．显然** Stieltjes **积分也与通常积分一样地具有加法性质． 此外，**Stieltjes **积分还具有以下的性质**:

(1) **若** $h(x) \ge 0$ **，则** $\int h(x)dG(x) \ge 0$ (**因为** $G(x)$ **递增**).

特别地，若 F(x) **是随机变量** x **的分布函数，则** $Eh(x) = \int h(x)dF(x)$ (**这在第一章中已经提到**).

(2) **若** $h_n(x) \ge 0$ **，且** $\sum_n h_n(x) < \infty$ **，则**

$$\int \sum_n h_n(x)dG(x) = \sum_n \int h_n(x)dG(x).$$

(3) 设右连续的递增函数列 $G_n(x)$ 满足 $\sum_k G_k(x) < \infty$，又 $h(x) \geq 0$ 则

$$\int h(x)d(\sum_k G_k(x)) = \sum_k \int h(x)dG_k(x).$$

注 事实上(2)和(3)给出了无穷和与积分的运算次序可以交换的条件．在(2)和(3)中，我们并未严格地给出 $h(x), h_n(x)$ 需要满足的条件，通常遇到的函数都可取成 $h(x), h_n(x)$．在本书中避免在数学上追求严格条件．

2. 更新过程的概念

2.1 作为 Poisson 过程推广的更新过程

更新现象是指数流的推广．也是一种按随机时刻到达的流，但是这个随机的流并不按独立同分布的指数分布随机变量的和到达，而是按非负独立同分布随机变量和的到达．

定义 4.1 假定 $\{T_k\}$ 是非负的独立同分布随机变量序列，且其二阶矩有限．记 $ET_1 = m$，$VarT_1 = s^2$，再假定 $a_0 \overset{\Delta}{=} P(T_1 = 0) < 1$．令

$$t_0 = 0, t_n = T_1 + \cdots + T_n, (n \geq 1), \quad N_t = \sup\{k : t_k \leq t\}, \tag{4.1}$$

即 N_t 是 $\{t_n\}$ 的计数过程（它们分别是 Poisson 过程与指数流的推广）．t_n 称为第 n 次更新时刻，随机序列 $\{t_n\}$ 称为更新流，T_n 称为第 n 个更新间隔，N_t 称为更新过程．

除了 Poisson 过程以外，一般的更新过程 N_t 就不是独立增量过程．但是更新过程与更新流之间仍由下面的关系联系起来

$$\{N_t \geq n\} = \{t_n \leq t\}. \tag{4.2}$$

定义 4.2 非负整值随机变量 h 称为关于随机变量列 $\{T_k\}$ 满足 Wald 条件，如果对于任意 n，事件 $\{h = n\}$ 与 $\{T_{n+1}, T_{n+2}, \cdots\}$ 独立．

引理 4.3 对于更新过程 N_t，在 t 固定时 $h \overset{D}{=} N_t + 1$ 关于 $\{T_k\}$ 满足 Wald 条件．

证明 $\{h = n\}$ 即随机事件 $\{N_t = n-1\}$，也就是 $\{t_{n-1} \leq t\} \setminus \{t_n \leq t\}$，后者可由随机变量组 $\{T_1, \cdots, T_n\}$ 完全地确定．故 $N_t + 1$（即 h）满足 Wald 条件． ?

注意 虽然 $N_t + 1$（即 h）满足 Wald 条件，但是 N_t 并不满足 Wald 条件．所以，对于 t_{N_t+1} 可以应用 Wald 等式(1.26)与(1.27)，而对于 t_{N_t} 则不可以应用 Wald 等式(1.26)与

(1．27))．

此外，即使在 N_t 是 Poisson 过程的情形，由于 $\{T_k\}$ 并不与 N_t 独立，作为随机多个项的独立和的 t_{N_t+1} 或 t_{N_t} 都不是复合 Poisson 过程．这样才能避免作不正确的推导.

命题 4．4　 $P(N_t=\infty)=0$．

证明　由大数定律, 我们有

$$P(N_t=\infty)=\lim_{n\to\infty}P(N_t\ge n)=\lim_{n\to\infty}P(t_n\le t)$$

$$=\lim_{n\to\infty}P(\frac{T_1+\cdots+T_n}{n}\le\frac{t}{n})=0. \qquad ?$$

记 t_n 的分布函数为 $F_n(t)=P(t_n\le t\}$．与 Poisson 过程类似地, 更新过程 N_t 的分布为

$$P(N_t=n)=F_n(t)-F_{n+1}(t). \qquad (4.\ 3)$$

假定 T_k 是第 k 次更新的部件的寿命, 利用 $F_n(t)$ 是时刻 t 前更新次数超过 n 的概率 $P(N_t\ge n)$，可以在此概率在容许小的控制条件下，设计部件备件的最小存储量. 也就是找尽可能小的 n，使 $F_n(t)$ 被控制在一个可以容许的“小“ 的范围内, 以达到减少储备量的目的. 但是 $F_n(t)=P(t_n\le t)$ 是 n 个与 T_1 独立同分布的随机变量的和的分布函数，除了极个别的例子外, 一般很不容易得到 $F_n(t)$ 的解析表达式. 然而, 在应用中可以借助于随机模拟得到它的近似. 例如, 多次 (例如 m 次) 生成 n 个独立 T_1 随机数并求和, 用此 m 个和中不超过 t 的频率, 作为 $F_n(t)$ 的估计值.

命题 4．5　更新过程 N_t 的期望函数称为更新函数. 它满足 $m(t)\overset{\Delta}{=}EN_t<\infty$，而且有

$$m(t)=\sum_{n=1}^{\infty}F_n(t) \qquad (4.\ 4)$$

证明　其有限性由 $P(T_1=0)<1$ 所保证, 我们略去其证明.　为了得到(4.4)式, 我们利用非负随机变量 $I_{\{t_n\le t\}}$，　它的求和与取期望可以交换, 所以

$$EN_t=E(\sum_{n=1}^{\infty}I_{\{t_n\le t\}})=\sum_{n=1}^{\infty}P(t_n\le t)=\text{右}.$$

命题 4.6　令$N_\infty \overset{\Delta}{=} \lim_{t\to\infty} N_t$，　则$P(N_\infty = \infty) = 1$.

证明　$P(N_\infty < \infty) \le P(\exists n,$ 使 $T_n = \infty) \le \sum_{n=1}^{\infty} P(T_n = \infty) = 0$.

1. 2　**更新函数的更新方程**

定理 4.7　设T_1的分布函数为$F_1(t)$，则更新函数$m(t)$满足积分方程

$$m(t) = F_1(t) + \int_0^t m(t-s)dF_1(s). \qquad (4.5)$$

证明　为了数学上更简单，我们假定分布函数F_1具有密度f_1. 于是F_n是n个独立同分布密度f的随机变量的和的分布. 因此，它的分布密度f_n满足关系

$$f_n = f_1 * f_{n-1}, f_2 = f_1 * f_1,$$

其中 $(f * g)(t) \overset{\Delta}{=} \int_0^t f(t-s)g(s)ds$，$f(t), g(t)$在$[0,\infty)$上定义，(易见$f * g = g * f$). 由

$$(\int_0^t f(t-s)F_n(s)ds\}' = (\int_0^t f(s)F_n(t-s)ds)' = \int_0^t f(s)f_n(t-s)ds = f_{n+1}(t)$$

可知

$$F_{n+1}(t) = \int_0^t f(t-s)F_n(s)ds = (f * F_n)(t) = (F_n * f)(t).$$

再用(4. 4)式便得

$$m(t) = F_1(t) + (F_1 + F_2 + \cdots) * f = F_1(t) + (m * f)(t). \qquad ?$$

注 1　更新函数$m(t)$是时刻t以前的平均更新次数EN_t，它是更新过程的一个极为重要的量，表达了部件的平均储备量. 方程(4. 5)可用来对$m(t)$作数值近似.　例如用典型的迭代方法，可以求得$m(t)$的数值近似解.　即令

$$m^{(0)}(t) = F_1(t), m^{(n+1)}(t) = F_1(t) + \int_0^t m^{(n)}(t-s)f(s)ds, (n \ge 0).$$

那么，在适当的条件下，就有$\sum_{k=0}^{n} m^{(k)}(t) \approx m(t)$.

注 2　设$h(t)$是一个在任意有限区间上有界的函数. 利用更新函数可以求解如下类型的积分方程

$$g(t) = h(t) + \int_0^t g(t-s)dF_1(s).$$

这种积分方程称为更新方程．更新方程在精算中的集体风险模型中，是一个最基本的数学工具．

下面我们来求解更新方程．假定更新间隔的分布函数$F_1(x)$具有密度$f_1(x)$．注意$F_n(x)$的密度函数f_n满足$f_{n+1}=f_1*f_n$，故而这时更新函数$m(t)$具有微商$m'(t)=\sum_{n=1}^{\infty} f_n(t)$．这样，更新方程可以改写为

$$g=h+(g*f_1)\text{，即}\, g*(1-f_1)=h.$$

由此得到更新方程的解

$$g=(1-f_1)^{-1}*h=(1+\sum_{n=1}^{\infty} f_n)*h=h+h*m'.$$

即

$$g(t)=h(t)+\int_0^t h(t-s)m'(s)ds.$$

如果更新间隔的分布函数$F_1(x)$不存在密度，那么只要利用 Stieltjes 积分的性质，就可以得到

$$g(t)=h(t)+\int_0^t h(t-s)dm(s).$$

这就是说，更新方程的解$g(t)$可以用更新函数$m(t)$表示．

注３　更新间隔的分布函数$F_1(x)$完全确定了更新函数$m(t)$．反之，更新函数$m(t)$也完全确定了更新间隔的分布函数$F_1(x)$．事实上它们可以通过 Laplace 变换相互表达．对于一个在$[0,\infty)$定义的非负函数$g(t)$，可以定义它的 Laplace 变换（记为$L_g(z)$）如下

$$L_g(z)=\int_0^{\infty} e^{-zt}g(t)dt.$$

当$g(t)$是更新时间T_1具有密度f_1时，其 Laplace 变换的概率含义为T_1的负指数矩，即$L_{f_1}(z)=Ee^{-zT_1}$．容易验证 Laplace 变换满足以下的乘法关系

$$L_{g*h}(z)=L_g(z)L_h(z).$$

在（４．５）式两边取 Laplace 变换，就得到其 Laplace 形式

$$L_m(z)=L_{F_1}(z)+L_{f_1}(z)L_m(z). \qquad (4.6)$$

由分部积分可以得到$L_{F_1}(z)=zL_{f_1}(z)$．代入（4．6）式就可以解出

$$L_m(z)=\frac{zL_{f_1}(z)}{1-L_{f_1}(z)}. \qquad (4.\ 7)$$

等价地有

$$L_{f_1}(z)=\frac{L_m(z)}{z+L_m(z)}. \qquad (4.\ 7)'$$

又由于非负随机变量的分布密度是由此随机变量的负指数矩唯一确定的，所以更新间隔的分布密度 f_1 可由更新函数$m(t)$唯一确定.

由(4. 7)式，只要用z的有理多项式来近似$L_m(z)$，再通过查一般函数的 Laplace 变换表，也可以得到$m(t)$的近似计算公式.

(如果更新间隔的分布函数$F_1(t)$不存在密度，那么这时候有 $Ee^{-zT_1}=\int_0^\infty e^{-zt}dF_1(t)$，这是分布函数$F_1(t)$的 Laplace-Stieltjies 变换，我们把它记为$L^S_{F_1}(z)$．再定义 $L^S_m(z)=\int_0^\infty e^{-zt}dm(t)$．与上面类似地可以得到

$$L^S_m(z)=\frac{L^S_{F_1}(z)}{1-L^S_{F_1}(z)}，\text{或}\ L^S_{F_1}(z)=\frac{L^S_m(z)}{1+L^S_m(z)} \qquad (4.\ 7)'')$$

1. 3　年龄与剩余寿命

定义 4．8　设t_n为更新流，N_t是其对应的更新过程．当t固定时，随机变量$a_t\overset{\Delta}{=}t-t_{N_t}$　称为（第N_t个更新元的）年龄；　而随机变量$g_t=t_{N_t+1}-t$　称为（第N_t个更新元的）剩余寿命.

命题 4．9（指数流的年龄与剩余寿命）设t_n是强度为l的指数流，则当t固定时有

（1）剩余寿命$g_t=t_{N_t+1}-t$服从强度为l的指数分布，即

$$g_t\overset{d}{=}T_1 \quad (\text{记号}\overset{d}{=}\text{表示两边的随机变量同分布}). \qquad (4.\ 8)$$

（2）年龄a_t的分布是如下的一个混合型的分布函数

$$P(a_t\le s)=I_{(0,t)}(s)(1-e^{-ls})+I_{[t,\infty)}(s). \qquad (4.\ 9)$$

即 a_t 与 $t \wedge t$ 同分布，其中 $t \wedge t \overset{\Delta}{=} \min(t, t)$，$t \sim \exp_l$.

证明 （1）随机变量 $g_t = t_{N_t+1} - t$ 分布函数为

$$P(g_t \le s) = P(t_{N_t+1} - t \le s)$$

$$= P(t_{N_t+1} \le t + s) = \sum_{n=0}^{\infty} P(N_t = n) P(t_{n+1} \le t + s \mid N_t = n)$$

$$= \sum_{n=0}^{\infty} P(N_t = n) P(N_s \ge 1) = P(N_s \ge 1) = 1 - e^{-ls}.$$

（2） 按定义当 $s \ge t$ 时 $P(a_t > s) = P(t - t_{N_t} > s) = 0$. 而当 $s < t$ 时有

$$P(a_t > s) = P(t - t_{N_t} > s) = \sum_{n=0}^{\infty} P(N_t = n, t_n < t - s)$$

$$= \sum_{n=0}^{\infty} P(N_{t-s} = n, N_t - N_s = 0) = \sum_{n=0}^{\infty} P(N_{t-s} = n) P(N_s = 0) = e^{-ls}.$$

推论４．１０ 指数流的平均更新年龄为:

$$Ea_t = \frac{1 - e^{-lt}}{l}. \tag{4. 10}$$

从而重新得到第 3 章 2.3 段中的结论：t 前最近更新时刻的数学期望为

$$Et_{N_t} = E(t - a_t) = \frac{e^{-lt} - 1 + lt}{l}.$$

证明 由命题４.９(2)可知，a_t 的分布函数在 t 点有一个跃度为 e^{-lt} 的跳跃，而在 $[0, t)$ 上有连续导数 $l \cdot e^{-ls}$，所以 a_t 的分布函数可以看成为如下的混合型分布

$$P(a_t \le s) = (1 - e^{-lt}) F_1(s) + e^{-lt} F_2(s),$$

其中 $F_1(s)$ 具有概率密度 $f_1(s) = \frac{l \cdot e^{-ls}}{1 - e^{-lt}} I_{[0,t)}(s)$，而 $F_2(s) = I_{[t,\infty)}(s)$. 由于这两个分布的数学期望分别为 $\frac{1 - l \cdot te^{-lt} - e^{-lt}}{l \cdot (1 - e^{-lt})}$ 与 t，因此 $Ea_t = \frac{1 - e^{-lt}}{l}$.

命题４.１１ （一般更新流的平均剩余寿命 ）对于一般更新过程的剩余寿命 g_t 有

$$Eg_t = (1 + m(t)) ET_1 - t. \tag{4. 11}$$

证明 由于$h\overset{\Delta}{=}N_t+1$满足 Wald 条件，我们仍可用 Wald 等式（参见(1. 26)式及(1. 27) 式)得到$Et_h=EhET_1$，从而有$Et_{N_t+1}=(1+m(t))ET_1$。

用同样的推理及$Var(t_h)=EhVar(T_1)+Var(h)ET_1^{\ 2}$，$Eh^2=\sum_{n,m=1}^{\infty}P(h>n,m)$可以得到（我们略去冗长的推演过程）

$$Eg_t^{\ 2}=(1+m(t))ET_1^{\ 2}-2(t+\int_0^t m(s)ds)ET_1+t^2\ . \qquad (4.\ 12)$$

我们还不加证明地指出下面的渐近关系

$$Eg_t\overset{t\to\infty}{\to}\frac{ET_1^{\ 2}}{2ET_1}\ ,\ Eg_t^{\ 2}\overset{t\to\infty}{\to}\frac{ET_1^{\ 3}}{3ET_1}, \qquad (4.13)$$

此关系可以为储存设计提供参考.

对t前的更新时刻t_{N_t}的分布，则有下面的积分表示

定理 4．12 （用更新间隔的分布函数$F_1(t)$表达t前的更新时刻t_{N_t}的分布函数）

$$P(t_{N_t}\le u)=(1-F_1(t))+\int_0^u(1-F_1(t-s))dm(s),(u\le t)\ . \qquad (4.\ 14)$$

证明 利用独立和t_n的 Markov 性质，我们得到

$$\text{左}=\sum_{n=0}^{\infty}P(t_n\le u,N_t=n)=\sum_{n=0}^{\infty}P(t_n\le u,t_{n+1}>t)$$

$$=(1-F_1(t))+\sum_{n=1}^{\infty}P(t_n\le u,t_{n+1}>t)$$

$$=(1-F_1(t))+\sum_{n=1}^{\infty}\int_0^u P(t_{n+1}>t\,|\,t_n=s)dF_n(s)$$

$$=(1-F_1(t))+\sum_{n=1}^{\infty}\int_0^u(1-F_1(t-s))dF_n(s)=\text{右}.$$

2 更新定理与更新次数的正态近似

在应用中，主要是要知道更新过程(更新次数)的渐近概率规律，以作为设计的参考.

2. 1 更新定理

定理 4．13 (更新定理) 对于更新过程有

(1) $P(\lim_{t\to\infty}\frac{N_t}{t}=\frac{1}{ET_1})=1$ （这里 ET_1可以取 $+\infty$).

(2) **基本更新定理**

$$\frac{m(t)}{t} \to \frac{1}{ET_1}, (t \to \infty) \quad (\text{这里 } ET_1 \text{ 可以取 } +\infty). \qquad (4.15)$$

(注 当 $Var(T_1) < \infty$ 时，还有

$$\frac{Var(N_t)}{t} \to \frac{Var(T_1)}{(ET_1)^3}, (t \to \infty). \qquad (4.16)$$)

证明 (1)利用 $r_{N_t} \le t < t_{N_t+1}$， 在 $t \to \infty$ 时，应用强大数定律便得 $\frac{t_{N_t}}{N_t} \xrightarrow{a.e.} ET_1$. (2)在直观上看就是对于(1)取数学期望，然而其严格证明较为复杂，本书略去.

注 在强度为 l 的指数流情形， 极限关系就简化为恒等关系，即对于任意 t 均有 $\frac{m(t)}{t} = \frac{1}{ET_1}$ ， $\frac{Var(N_t)}{t} = \frac{Var(T_1)}{(ET_1)^3}$ ， 这是因为 $ET_1 = \frac{1}{l}$ ， $Var(T_1) = \frac{1}{l^2}$ ， $Var(N_t) = m(t) = l \cdot t$.

*2. 2 更新过程的正态近似

对于更新过 N_t 而言，事件 $\{N_t \ge n\}$ 即 $\{t_n \le t\}$ ，而 后者作为独立同分布的随机变量的和的分布近似正态，所以 ，更新过程在时间发展充分长后，其分布就近似正态。设 $ET_1 = m$ ， $Var(T_1) = s^2 < \infty$ ，则定理 4.13 说明当 $t \to \infty$ 时有 $EN_t \approx \frac{t}{m}$ ， $Var(N_t) \approx \frac{ts^2}{m^3}$ ，于是我们有

定理 4．14 (正态近似定理) 若 $Var(T_1) = s^2 < \infty, ET_1 = m$ ，则

$$\frac{N_t - \frac{t}{m}}{s\sqrt{\frac{t}{m^3}}} \approx N(0,1), (t \to \infty). \qquad (4.17)$$

证明 $P(\frac{N_t - \frac{t}{m}}{s\sqrt{\frac{t}{m^3}}} \le x) = P(N_t \le xs\sqrt{\frac{t}{m^3}} + \frac{t}{m})$. 令 $b(t)$ 是大于 $xs\sqrt{\frac{t}{m^3}} + \frac{t}{m}$ 的最小整数，那么等式的右边等于

$$P(N_t \le b(t)) = P(t_{b(t)} \ge t) = P(\frac{t_{b(t)} - mb(t)}{s\sqrt{b(t)}} \ge \frac{t - mb(t)}{s\sqrt{b(t)}})$$

$$\approx P(\frac{t_{b(t)} - mb(t)}{s\sqrt{b(t)}} \geq -x) \xrightarrow{t\to\infty} 1-\Phi(-x) = \Phi(x),$$

其中$\Phi(x)$是$N(0,1)$的分布函数. 这里用了t_n服从中心极限定理以及由$b(t)$的定义所引出的近似关系

$\frac{t - mb(t)}{s\sqrt{b(t)}} \approx -x$.

注 在强度为l的指数流的情形，此定理即是中心极限定理$\frac{N_t - l\cdot t}{\sqrt{l\cdot t}} \approx N(0,1), (t\to\infty)$，因为此时有$m = \frac{1}{l}, s^2 = \frac{1}{l^2}$.

2. 3 Blackwell 定理与主更新定理

在本段中，我们不再假定更新间隔T_1有密度函数，它可以是离散的随机变量.

定义 4．15 (格点分布) 设随机变量x只取某个正数 (更常见的是正整数)d的整数倍，而且这个d是满足此性质的最大者(d称为周期)，则称此随机变量具有$d-$格点分布.

我们不加证明地介绍两个在更新理论中最重要的定理，Blackwell 定理及作为其延伸的主更新定理 (key renewal theorem)

定理 4．16 (Blackwell 定理) 设更新间隔T_1的数学期望$ET_1 = m$.

(1) 若更新间隔T_1不是格点分布，则

$$\lim_{t\to\infty}[m(t+s) - m(t)] = \frac{s}{m}; \tag{4．18}$$

(2) 若更新间隔T_1是$d-$格点分布，则

$$\lim_{n\to\infty} m(nd) = \frac{d}{m}. \tag{4．19}$$

(注 (1)的形式可由定理 4.13 猜得，其严格证明的主要点在于极限的存在性，一旦证明存在性，那么此极限作为s的函数易见是可加的，等式就自然成立. 本定理的证明可参见文献 [H]).

定理 4．17 (主更新定理) 若更新间隔T_1不是格点分布，$ET_1 = m$，则对可积函数g有

$$\lim_{t\to\infty}\int_0^t g(t-s)dm(s) = \int_0^\infty \frac{g(t)}{m}dt. \tag{4．20}$$

注 1 在指数流时(4．20)是显然的.

注 2 Blackwell 定理中的(1)，相当于在主更新定理中取$g(t) = I_{(t,t+s]}$. 事实上，由

严格的概率理论，可以证明 Blackwell 定理的(1)与主更新定理是等价的.

主更新定理中左边的积分经常出现，例如，在更新方程的解中．而主更新定理的主要应用就是用以求更新方程解的渐近表示.特别地，在计算与更新过程有关的某个数学期望在t很大时的渐近值时，常常先以首次到达某处的不同时间作为条件，用全概率公式(或全期望公式)得到一个更新方程，在求得此更新方程的解后，再用主更新定理便可得到近似的数学期望值.

2．4　更新间隔为正整值随机变量的更新过程

对于这种更新流，其更新时刻$t_k(k>0)$的取值都是自然数,因而是格点分布．此时的更新过程N_t是更新序列N_n，而$t_0=0, t_k=T_1+\cdots+T_k, (k\ge 1)$，$N_n=\sup\{k: t_k=n\}$.

由于T_1是离散的随机变量，在计算更新函数(时刻n的平均更新次数) $m(n)\overset{\Delta}{=}EN_n$ 时，递推关系(4. 5)式应该作相应的改变．记

$$\mathrm{p}^T=(p_1,\cdots,p_k,\cdots),\ p_k=P(T_1=k),\ \mathrm{p}^{m}=\mathrm{p}*\mathrm{p}^{m-1},\quad \mathrm{p}^2=\mathrm{p}*\mathrm{p},$$

其中 $(\mathrm{p}*\mathrm{q})_k=\sum_i p_i q_{k-i}$ 是序列的卷积运算．再记平均更新次数列组成的无穷维向量为

$\mathrm{m}^T=(m(1),\cdots,m(k),\cdots)$，$T_1$的分布函数取值的序列为 $\mathrm{F}_1^T=(F_1(1),\cdots,F_1(k),\cdots)$．那么,(4. 4)式与(4. 5)式就分别成为

$$\mathrm{m}=\sum_{k=1}^{\infty}\mathrm{p}^k \tag{4. 4)'$$

$$\mathrm{m}=\mathrm{F}_1+\mathrm{m}*\mathrm{p}\ . \tag{4. 5)'$$

于是一切相应的结论都类似地成立.

3 更新过程的变种模型

3. 1 交错更新过程

定义 4．18　假定更新间隔流$\{T_n\}$可以分前后两个独立的阶段，即$T_n=T_n'+T_n''$,且T_n'在前，T_n''在后，彼此交错地到来．再假定对于不同的n，(T_n',T_n'')是二维的独立同分布随机向量序列．这种数学模型称为交错更新过程.

交错更新过程的典型例子是随机开关，其中“开间隔”与“关间隔”各自独立同分布，并且彼此交错.

对于交错更新过程，可以证明如下的比例极限定理:

定理4．19　(比例极限定理)

如果更新间隔流$\{T_n'\}$与$\{T_n''\}$都不是格点分布，那么，在时间趋于∞时，交错更新过程所描述的系统处于“T’状态”与“T’’状态”的概率及总时间所占的比例的极限，都与初始状态无关，且分别以ET_1与ET_1'的比例分配，即

$$\lim_{t\to\infty} P(\text{在}t\text{时刻，系统处于}T'\text{状态}) = \frac{ET_1'}{E(T_1'+T_1'')}, \qquad (4.\ 21)$$

$$\lim_{t\to\infty} \frac{[0,t]\text{中系统处于}T'\text{状态的时间}}{t} = \frac{ET_1'}{E(T_1'+T_1'')}. \qquad (4.\ 22)$$

这两个事实非常符合直观，我们略去它的证明.

3. 2　延迟更新过程

如果允许第一个更新间隔与其后的更新间隔的分布不同，就称为延迟更新过程．为了易于区别，我们常常把第一个更新间隔记为T_0，而把其后的更新间隔记为$T_1, T_2, \cdots$．于是$\{T_n, n \ge 1)$是独立同分布的随机变量列，且它与T_0独立.

延迟更新过程的典型例子有:

（1）对于固定的时刻s，更新过程$\{N_t : t \ge 0)$在时间s后的更新情况，即更新间隔流为$\tilde{T}_n = T_{N_s+n+1}, (n \ge 1), \tilde{T}_0 = t_{N_s+1} - s$．直接计算可以得到(从直观上也可看出)$\tilde{T}_n (n \ge 1)$与$T_1$同分布，而$\tilde{T}_0$是$s$后的剩余寿命，与$T_1$的分布显然不同．所以$\{\tilde{T}_n : n \ge 0\}$是延迟更新流，其计数过程$\tilde{N}_t$就是延迟更新过程.

（2）在下一章中的时齐 Markov 链，在时刻t前对于一个固定的常返态的返回次数是一个延迟更新过程.

在直观上我们很容易接受下面的结论:

对于延迟更新过程而言，更新定理， Blackwell 定理和主更新定理的形式仍然不变，且其更新函数$m(t)$仍然满足 (唯一不同处为，T_0的分布函数$F_0(t)$与T_1的分布函数$F_1(t)$不同)

$$m(t) = F_0(t) + \int_0^t m(t-s) dF_1(s).$$

3. 3　带酬更新过程

如果对于每次更新都支付独立同分布的随机酬金(修理费或理赔金)，这样的更新过程就称为带酬更新过程．更确切一些，记第 n 次随机酬金为X_n，并假定$\{X_n\}$是独立同分布的随机变量列，且与更新过程$\{N_t : t \ge 0\}$独立．再记t前的累计酬金(累计修理费或累计理赔金)为R_t．那么，R_t称为酬金过程．而二维向量随机过程$\{N_t, R_t\}(t \ge 0)$ 则称为带酬更新过程.

带酬更新过程的酬金过程可表达为 $R_t \overset{\Delta}{=} \sum_{n=1}^{N_t} X_n$．当更新过程是 Poisson 过程时， 带酬更新过程的酬金过程就是复合 Poisson 过程．所以，作为数学模型的带酬更新过程的酬金过程，是复合 Poisson 过程的推广和深化．如果某种保险项目的事故发生流不是指数流，那么t 前累计理赔额就是一个带酬更新过程的酬金过程.

定理 4．20（酬金过程的更新定理） 对于酬金过程，只要数学期望 $ET_1 = m < \infty$，$ER_1 = g < \infty$，就有

(1) $$P(\lim_{t\to\infty} \frac{R_t}{t} = \frac{ER_1}{ET_1}) = 1 . \tag{4. 23}$$

(2) 酬金过程的基本更新定理 $$\frac{ER_t}{t} \to \frac{ER_1}{ET_1}, (t \to \infty) . \tag{4. 24}$$

证明提示 (1)的证明利用 $r_{N_t} \le t < t_{N_t+1}$， 在 $t \to \infty$ 时利用强大数定律便可得

$$\frac{t_{N_t}}{N_t} \xrightarrow{a.e.} ET_1 , \quad \frac{R_{N_t}}{N_t} \xrightarrow{a.e.} ER_1 ，从而. \frac{R_{N_t}}{t_{N_t}} \xrightarrow{a.e.} \frac{ER_1}{ET_1} .$$

同样, (2)在直观上看是简单地对(1)取数学期望. 但是严格证明较为复杂，从略.

酬金过程的正态近似为

$$\frac{R_t - \frac{ER_1}{ET_1} t}{\sqrt{\frac{t}{ET_1} E(R_1 - \frac{ER_1}{ET_1} T_1)^2}} \xrightarrow{t\to\infty} N(0,1) . \tag{4. 25}$$

证明提示 在 $t \to \infty$ 时，分别用 $R_1 + \cdots + R_{N_t} + o(1)$ 与 $T_1 + \cdots + T_{N_t} + o(1)$ 代替 R_t 与 t，再利用以下的独立和 $(R_1 + \frac{ER_1}{ET_1} T_1) + \cdots + (R_{N_t} + \frac{ER_1}{ET_1} T_{N_t})$ 的中心极限定理，即可证明.

由于更新过程一般并不是独立增量过程，所以带酬更新过程的酬金过程也不是独立增量过程了. 使用这种模型作为保险中的集体风险的数学模型，无论是在带利率或不带利率这两种情形，时刻 t 前不破产概率及破产前后保险公司的盈余等等的分析，都将会扩大模型的有效使用范围.

4 再生过程与其相系的更新过程

4. 1 再生过程的概念

定义 4．21 随机过程 X_t 称为再生过程 (regenerative process)，如果存在有限值(X_t)停时T，使随机过程$\{X_t : t \ge 0\}$与随机过程$\{X_{T+t} : t \ge 0\}$同分布 (即有相同的有限维分布族)，而且$\{X_{T+t} : t \ge 0\}$与T 独立. 此时随机变量T 则称为一个再生时刻.

4.2 与再生过程相系的更新过程

对于再生过程 X_t，存在再生时刻 T_1，使随机过程 $X_t^{(1)} \stackrel{\Delta}{=} X_{T_1+t}$ 与随机过程 X_t 同分布，

且$(X_t^{(1)})$与T_1独立. 于是$(X_t^{(1)})$也是再生过程. 由于它与(X_t)同分布,从而存在与T_1同分布的再现时刻T_2，使随机过程 $X_t^{(2)} \overset{\Delta}{=} X_{T_2+t}^{(2)}$ 与随机过程 $X_t^{(1)}$ 同分布，而且随机过程$(X_t^{(2)})$与T_2独立，于是T_2与T_1独立 (独立性的严格证明常较为复杂，但是直观上容易判定)，如此继续，就可以得到独立同分布的非负随机变量列$\{T_n\}$. 这就构成了一个产生更新过程的流，与它对应就有一个更新过程N_t.

T_n称为再生过程X_t的第n个循环时间，$[t_{n-1}, t_n)$称为再生过程X_t的第n个循环时段.

再生过程在结构上比更新过程更为基本. 一个再生过程可以有许多个与它相系的更新流,从而可以对应不同的更新过程. (例如在第 5 章中的 Markov 链回访每个固定常返状态的时间列，就是一个产生更新过程的流. 而在第 12 章中的一维扩散过程, 从每个固定状态出发，经过给定的第二个状态后再回访原状态的时间列，也是一个产生更新过程的流).

4. 3 比例极限定理在再生过程中的应用

比例极限定理在再生过程中的应用，主要有

(1). 用与再生过程相系的更新过程的各种更新定理求极限;

(2). 把一个循环时段划分为前后两个各自独立的两个时段，得到交错更新过程，再用比例极限定理求一些渐近分布

例４．２２ (更新过程中年龄的渐近分布)

设N_t更新间隔为$\{T_n\}$的更新过程. 它在t时刻所在部件的年龄为a_t. 则a_t是一个以T_n为第n个循环时段的再现过程. 作循环时间的如下分解:

$$T_n = T_n' + T_n'',\ T_n' = T_n \wedge s,\ \ T_n'' = T_n - T_n'.$$

那么由比例极限定理立刻推出年龄的如下的极限分布:

$$P(a_t \le s\} = P(\text{系统在时刻}t\text{处于}T_n'\text{状态}) \overset{t\to\infty}{\to} \frac{E(T_1 \wedge s)}{ET_1}.$$

例４．２３ (更新过程中剩余寿命的渐近分布)

设更新间隔为$\{T_n\}$的更新过程，在t时刻所在部件的剩余寿命为g_t. 则g_t也是一个以T_n为第n个循环时段的再现过程. 作循环时间的如下分解:

$$T_n = T_n' + T_n'',\ T_n'' = T_n \wedge s,\ \ T_n' = T_n - T_n''.$$

那么由比例极限定理立刻推出剩余寿命的如下的极限分布:

$$P(g_t \le s\} = P(\text{系统在时刻}t\text{处于}T_n''\text{状态}) \overset{t\to\infty}{\to} \frac{E(T_1 \wedge s)}{ET_1}.$$

推论４．２４ 任何更新过程中的年龄与剩余寿命，在$t \to \infty$时，具有相同的极限分布.

这个推论在排队理论中具有重要的意义，因为直观地看，如果忽略起点而倒向地看时间,那么剩余寿命就变成了年龄，而年龄就变为剩余寿命. 这是一种在时间倒逆下的对偶关系.

那么，这个推论的含义为：更新过程具有某种渐近的时间可逆性．于是人们就可以利用这种渐近的时间可逆性，把求一个渐近分布，化简为求另一个相当于它的对偶的渐近分布.

比例极限定理还可以推广为

定理４．２５　对于再现过程 X_t 及循环时段 T_1，只要 $E\int_0^{T_1}|f(X_t)|dt<\infty$，就有

$$\lim_{t\to\infty} f(X_t)=\frac{E(\int_0^{T_1} f(X_s)ds)}{ET_1}. \qquad (4.26)$$

4.4 存储模型的一个例子

例４．２６　(商品库存问题)

设某店经营一种商品，初始进货量为 S． 固定取一个值 $S_0<S$ 后，该店确定其进货策略为（称为 (S_0,S) 策略）：即当且仅当存货量少于 S_0 时进货，并使存货量补足到 S．假定顾客到达的间隔为非负的独立同分布的随机变量列 $\{x_n\}$，而他们的需求量为与之独立的独立同分布的随机变量列 $\{Z_n\}$，且 Z_n 具有非格点的分布函数 F_Z． 该店在时刻 t 的存货量记为 X_t．令 T_1 为开张后第一次进货时刻，T_2 为第二次进货时刻，依此下去．于是 X_t 是再生过程，而 $\{T_n\}$ 就构成了更新流.

这个例子概括了存储模型的基本特征．这里有包括

(1) 在 $(0,t]$ 中到达顾客的人数是一个以 $\{x_n\}$ 为更新间隔流的更新过程，记为 N_t．

(2) 设第 n 个顾客需要购置的商品数为 Z_n，它们是与之独立的独立同分布的随机变量列．于是需求流 $\{Z_n\}$ 也是一个更新间隔流，它又对应于另一个更新过程，记为 ${}^{(Z)}N_t$，这是一个描述储备消耗的更新过程．(而在时刻 t 的累计需求量，记为 W_t，则可以看成为一个"酬金"为 $\{Z_n\}$ 的带酬更新过程，其"酬金过程"为 $W_t\overset{\Delta}{=}\sum_{n=1}^{N_t}Z_n$).

(3) 购进商品的流(商品补给流)，它相当于一个在每个更新间隔 T_n 开始时，带上常数 "酬金" $S-S_0$ 的带酬更新过程.

存储模型的首要问题是研究储量过程 X_t．它涉及三个更新过程，其中前两个(顾客流与需求流)彼此独立，但是第三个(商品补给流)本质地依赖于前两个．所以储量过程是一个远比更新过程复杂的数学模型．(事实上，它可以纳入第 7 章排队模型的框架之中，但它是一个较

为复杂的排队模型).

商品补给发生在累计售出量达到$S-S_0$的时刻，第 1 次补给发生的时刻的顾客人数为:

$$t_{S-S_0} \overset{\Delta}{=} \min\{n: Z_1+\cdots+Z_n > S-S_0\} = {}^{(Z)}N_{S-S_0}+1,$$

从而 $T_1 = x_1+\cdots+x_{t_{S-S_0}}$.

下面我们研究商品储量过程 X_t 的渐近分布，以备进货时参考．为此我们注意，一个再生循环T_n开始于商品存量补给为 S 的时刻，而结束于商品存量减少到 S_0的时刻．而存量减少到u的时刻，恰好是在此循环时段中，累计售出量达到$S-u$的时刻，此时的顾客人数为:

$$t_{S-u} \overset{\Delta}{=} \min\{n: Z_1+\cdots+Z_n > S-u\} = {}^{(Z)}N_{S-u}+1.$$

对于固定的正整数u，令循环T_n中在第 $N_{s_{S-u}}$ 个顾客到达的时刻之前的一段为$T_n{}'$，余下的一段为$T_n{}''$．直观上看$\{T_n{}', T_n{}''\}$是独立同分布的二维随机向量序列，这样就得到了一个交错更新过程．用比例极限定理及 Wald 引理，我们便得到

$$\lim_{t\to\infty} P(X_t \ge u) = \lim_{t\to\infty} P(\text{系统在}t\text{时刻处于}T_n{}'\text{状态}\}$$

$$= \frac{ET_1{}'}{ET_1} = \frac{E(x_1+\cdots+x_{t_{S-u}})}{E(x_1+\cdots+x_{t_{S-S_0}})} = \frac{Et_{S-u}}{Et_{S-S_0}}$$

$$= \frac{E({}^{(Z)}N_{S-u})+1}{E({}^{(Z)}N_{S-S_0})+1} \overset{\Delta}{=} \frac{m_{F_Z}(S-u)+1}{m_{F_Z}(S-S_0)+1},$$

这里更新过程${}^{(Z)}N_t$的更新函数$m_{F_Z}(t)$可以用 1. 2 段的方法通过近似计算得到.

* §5. Erlang 更新过程

5. 1 Erlang更新过程的定义

定义 4．27 k个独立同指数分布的随机变量的和的分布(即是$\Gamma(k, l)$分布)，在排队理论中称为k阶 Erlang 分布，其密度函数为

$$f_k(t) \overset{\Delta}{=} l e^{-lt} \frac{(l\cdot t)^{k-1}}{(k-1)!} I_{[0,\infty)}(t).$$

若更新间隔流$\{T_n\}$的分布为$\Gamma(k, l)$，则此更新流称为k阶 Erlang 流，其对应的更新过程N_t称为 k 阶 Erlang 更新过程．

注 用 Erlang 流作为更新元件的寿命流的数学模型要比指数流(也称 Poisson 流)更为合理，因为后者假定更新间隔服从指数分布，因而有″无记忆性″，即

$$P(T_i > t+s \mid T_1 > s) = P(T_1 > t),$$

也就是更新间隔的概率分布与它的过去的情形(s 前的情形)无关．如果把设备的自然更新(损坏)作为更新间隔，那么，指数分布的假定就是设备没有折旧，这显然不合理的．诚然,由于指数流的计算远比其它更新流简单，所以在某些情况下，我们还用指数流作为设备寿命模型的粗略近似．当然，我们总希望采用别的比较合理但相对地还容易计算其统计特征的更新流，其中最简单的就是 Erlang 流．它的优点有二:

（1） 在$[0,\infty)$上具有密度的随机变量的分布函数，都可以用 Erlang 分布的分布函数的混合来近似．

(这说明了在排队系统中，用混合的 Erlang 流也可以足够好地近似一般的更新流)．

（2） Erlang 流的更新过程的母函数可以明显的写出来（特别是 2 阶 Erlang 流），所以成为人们选用的对象．

关于（1），我们有以下定理

定理４．２８ （用混合 Erlang 分布近似正随机变量的分布(参阅［T］)

设 $F(t)$ 是$[0,\infty)$ 上的一个分布函数。对于 $h>0$ 定义

$$F_h(t) = \sum_{n=1}^{\infty} [F(nh) - F((n-1)h](1 - \sum_{k=0}^{n-1} \frac{(\frac{t}{h})^k}{k!} e^{-\frac{t}{h}}),$$

即 $F_h(t)$ 是无穷多个权重分别为 $p_n = F(nh) - F((n-1)h)$ 的 $Erlang(n, \frac{1}{h})$ 分布的混合分布的分布函数，那么 $F_h \xrightarrow{d} F$，即在 F 的任意连续点 t，都有

$$F_h(t) \xrightarrow{d} F(t) \qquad (h \to 0).$$

5. 2 **Erlang 更新过程的矩母函数**

我们来推导 k 阶 Erlang 更新过程 N_t 的矩母函数

$$G_k(t,z) \overset{\Delta}{=} Ez^{N_t} = \sum_{n=0}^{\infty} P(N_t = n) z^n$$

$$= \sum_{n=0}^{\infty} P(t_n \le t < t_{n+1}) z^n,\ (t_0 = 0, t_n = T_1 + \cdots + T_n)$$

$$= \sum_{n=0}^{\infty} [P(t_n \le t) - P(t_{n+1} \le t)] z^n$$

$$= P(t_0 \le t) + z\sum_{n=1}^{\infty} P(t_n \le t) z^{n-1} - \sum_{n=1}^{\infty} P(t_n \le t) z^{n-1}$$

$$= 1 + (z-1)\sum_{n=1}^{\infty} P(t_n \le t) z^{n-1}.$$

由于$T_n \sim \Gamma(k,l)$，便有 $t_n \sim \Gamma(kn,l)$．令 $z = y^k$．我们有

$$\sum_{n=1}^{\infty} P(t_n \le t) z^{n-1} = \sum_{n=1}^{\infty} \left(\int_0^t \frac{(lu)^{kn-1}}{(kn-1)!} lu^{-lu} du\right) y^{k(n-1)}$$

$$= y^{1-k} \sum_{n=1}^{\infty} \left(\int_0^t \frac{(luy)^{kn-1}}{(kn-1)!} lu^{-lu} du\right)$$

$$= y^{1-k} \int_0^t l e^{-lu} \sum_{m+1=n,2n,\cdots} \frac{(lyu)^m}{m!} du$$

记 k 阶原单位根为 $e_k \overset{\Delta}{=} e^{i\frac{2p}{k}}$，(即方程 $x^k = 1$ 的 k 个根为 $(e_k)^j, j = 1,2,\cdots k)$．那么

$$\sum_{j=0}^{k-1} (e_k)^{jl} = \sum_{j=0}^{k-1} e^{i\frac{2p}{k}jl} = \begin{cases} k & (l\text{是}k\text{的倍数}) \\ 0 & (l\text{非}k\text{的倍数}) \end{cases}.$$

所以

$$\sum_{m+1=n,2n,\cdots} \frac{x^m}{m!} = \sum_{m=0}^{\infty} \frac{1}{k} \sum_{j=0}^{k-1} (e_k)^{j(m+1)} \frac{x^m}{m!} = \frac{1}{k} \sum_{j=0}^{k-1} (e_k)^j e^{x(e_k)^j}.$$

从而

$$\sum_{n=1}^{\infty} P(t_n \le t) z^{n-1} = y^{1-k} \int_0^t l e^{-lu} \frac{1}{k} \sum_{j=0}^{k-1} (e_k)^j e^{lyu(e_k)^j} du$$

$$= y^{1-k} \frac{1}{k} \sum_{j=0}^{k-1} \frac{(e_k)^j}{1 - y(e_k)^j} (1 - e^{-lt(1-y(e_k)^j)}).$$

再注意到 $z = y^k$，我们就得到下面的定理

定理４．２９　k 阶 Erlang 更新过程 N_t 的矩母函数为

$$G_k(t,z) = 1 + \frac{1}{k}(z-1) z^{\frac{1}{k}-1} \sum_{j=0}^{k-1} \frac{(e_k)^j}{1 - z^{\frac{1}{k}}(e_k)^j} (1 - e^{-lt(1-z^{\frac{1}{k}}(e_k)^j)}).$$

特别地，2 阶 Erlang 流的更新过程 N_t 的矩母函数为

$$G_2(t,z) = e^{-lt}[ch(l \cdot t\sqrt{z}) + \frac{1}{\sqrt{z}} sh(l \cdot t\sqrt{z})].$$

从而有

$$EN_t (= \frac{\partial G_2(t,z)}{\partial z}|_{z=1}) = \frac{l \cdot t}{2} + \frac{1}{4}(e^{-2lt} - 1),$$

$$Var(N_t) (= \frac{\partial^2 G_2(t,z)}{\partial z^2}|_{z=1} + EN_t - (EN_t)^2)$$

$$=\frac{l\cdot t}{4}-\frac{l\cdot t}{2}e^{-2lt}+\frac{1}{4}e^{-lt}sh(l\cdot t)-\frac{1}{4}e^{-2lt}sh^2(l\cdot t).$$

习题4

1. 某仪器受到一个强度为 l 的 Poisson 流的振动. 假定每次振动的损失为独立同分布的随机变量列 $\{L_i\}$，并且损失随时间衰减，即第 i 次振动发生的时刻 t_i 后的时间 t 的损失为 $L_i e^{-a(t-t_i)}$. 设 t 前振动的次数为 N_t. 那么在时刻 t 的总损失为 $L(t)=\sum_{i=1}^{N_t}L_i e^{-a(t-t_i)}$. 求证 $E[L(t)]=l\cdot EL_1\cdot\frac{1-e^{-at}}{a}$.

2. 证明更新过程 $\{N_t, t\ge 0\}$ 为 Poisson 过程，当且仅当它同时满足以下条件

   (1) t 固定时 $N_t \sim Poisson_{lt}$;

   (2) 其更新流 $\{t_k\}$ 对于任意 n，$(t_1,\cdots,t_n\}$ 在条件概率 $P(*\mid N_t=n)$ 下的分布与 $(U_1^*[0,t],\cdots,U_n^*[0,t])$ 的分布相同，其中 $U_1^*[0,t],\cdots,U_n^*[0,t]$ 为独立同分布的 $[0,t]$ 均匀随机变量 $x_1,\cdots,x_n$ 按从下到大排序后的随机变量(即次序统计量)的联合分布.

3. 如果 N_t 是更新间隔为 $T_n\sim\begin{pmatrix}1 & 2\\ 1-p & p\end{pmatrix}$ 的更新过程. 分别求 N_1 和 N_3 的分布.

4. 用 Laplace 变换证明更新过程为 Poisson 过程的充要条件是它的更新函数为 $m(t)=l\cdot t$.

5. 设 t_n 为更新过程 N_t 的第 n 次更新时刻，证明 N_t 是 Poisson 过程的充要条件是对于 $[0,\infty)$ 上的任意非负连续函数 $f(t)$，都有

$$E[\sum_{n=1}^{\infty} f(t_n)]=l\int_0^{\infty} f(t)dt.$$

6. 对于 Poisson 过程 N_t， 求(1) 时刻 t 前的年龄 a_t 与时刻 t 后的余寿 g_t 的联合分布.

   (2) $P(g_t>u\mid a_{t+u}>s)$，$P(g_t>u\mid a_t=s)$.

7. 对于更新过程 N_t， (1) 求 $E[T_1+\cdots+T_{N_t}\mid N_t=n]$，问它是否等于 $E(T_1+\cdots+T_n)$?

   (2) 证明 $E[\frac{T_1+\cdots+T_{N_t}}{N_t}\mid N_t>0]=E(T_1\mid T_1<0)$.

8. 如果 N_t 是更新间隔为 T_n 为延迟指数分布(即其密度为 $f(t)=Ce^{-l(t-d)}I_{(t>d)}$, $(d>0)$.

求 $P(N_t \geq n)$.

9. 说明对于固定的 t , 强度为 l 的 Poisson 过程的 t 前寿命 a_t 与 t 后剩余寿命 g_t 是相互独立的随机变量,

并且满足 $E(a_t + g_t) = \frac{2}{l}$.

10. 如果 N_t 是 Erlang(2)更新过程, 即其更新间隔为 T_n 具有分布密度 $f(t) = l^2 t e^{-lt} I_{[0,\infty)}(t)$, 求其更新

函数 $m(t)$.

11. 对于更新间隔流 $\{T_n\}$, 记 t 前最近更新时刻为 $Y_t = t_{N_t}$. 证明

$$\lim_{t\to\infty} P(Y_t \in A) = \frac{E[\int_0^{T_1} I_A(Y_s)ds]}{ET_1}.$$

(提示: 用比例极限定理).

12. 设交错更新过程中, 更新间隔 $T_n' \sim \exp_l, T_n'' \sim \exp_m$, 把系统处于此两段更新间隔中, 分别称为处

于 "开" 状态和 "闭" 状态. 证明: P(系统在 t 时刻处于开状态) $= \frac{m}{l+m}(1 - e^{-(l+m)t})$.

龚光鲁，钱敏平著 应用随机过程教程及其在算法与智能计算中的应用

清华大学出版社,2003

第 5 章 时间离散的 Markov 链

1 Markov 链的概念

1.1 定义与 Markov 性质

定义5.1 随机过程可能取到的值(状态)组成的集合记为S，称为随机过程的状态空间．随机序列$\{x_n : n \geq 0\}$称为 Markov 链, 如果这些随机变量都是离散的 (状态空间S至多是一个可数集, 即或者S是有限集, 或者S可与自然数一一对应), 而且对于 $\forall n \geq 0$ 及任意状态$i, j, i_0, \cdots, i_{n-1}$, 都有

$$P(x_{n+1} = j \mid x_n = i, x_{n-1} = i_{n-1}, \cdots, x_0 = i_0) = P(x_{n+1} = j \mid x_n = i) . \quad (5.1)$$

这个性质称为 Markov 性质.

Markov 性的等价性质 1

若A为任意形如$\{x_0(w) = i_0, x_1(w) = i_1, \cdots., x_{n-1}(w) = i_{n-1}\}$的事件的并，即

$$A = \bigcup_k \{x_0 = i_{0,k}, x_1 = i_{1,k}, \cdots, x_{n-1} = i_{n-1,k}\}$$

则有

$$P(x_{n+1} = j \mid A, x_n = i) = P(x_{n+1} = j \mid x_n = i) . \quad (5.2)$$

Markov 性的等价性质 2

进一步还可以证明 Markov 性等价于：对于过程在时刻 n+1 及其以后的时刻所确定的事件B及等价性质中之A有

$$P(B \mid A, x_n = i) = P(B \mid x_n = i) . \quad (5.3)$$

或

$$P(AB \mid x_n = i) = P(A \mid x_n = i) P(B \mid x_n = i) \quad (5.3)'$$

(5．3) 式的含意为: 如果过程在时刻 n 处于状态i，那么不管它以前处于什么状态，过程以后处于什么状态的概率是一样的. 这就说明了，Markov 链在已知‘现在’的条件下，“将来”与“过去”是条件独立的．我们把它另列为一条性质。

Markov 性的等价性质 3

在已知“现在”的条件下，“将来”与“过去” 是条件独立的.

Markov 性的等价性质 4

对 Markov 链$\{x_n\}$及$\forall m \geq 1, n \geq 0$及任意状态$i, j, i_0, \cdots, i_{n-1}$, 有

$$P(x_{n+m} = j \mid x_n = i, x_{n-1} = i_{n-1}, \cdots, x_0 = i_0) = P(x_{n+m} = j \mid x_n = i) . \quad (5.4)$$

证明 对m作归纳法. m=1 **时即为**(5. 1)**式**. **设** m **时**(5. 4)**式正确**, **今证** m+1 **时也正确**.

$$P(x_{n+m+1}=j\mid x_n=i,x_{n-1}=i_{n-1},\cdots,x_0=i_0)$$

$$=\sum_k P(x_{n+m+1}=j,x_{n+1}=k\mid x_n=i,x_{n-1}=i_{n-1},\cdots,x_0=i_0)$$

$$=\sum_k P(x_{n+m+1}=j\mid x_{n+1}=k,x_n=i,x_{n-1}=i_{n-1},\cdots,x_0=i_0)\times$$

$$P(x_{n+1}=k\mid x_n=i,x_{n-1}=i_{n-1},\cdots,x_0=i_0).$$

利用归纳法假设及(5. 1)**式**, **上式简化为**

$$=\sum_k P(x_{n+m+1}=j\mid x_{n+1}=k)\ P(x_{n+1}=k\mid x_n=i)$$

$$=\sum_k P(x_{n+m+1}=j\mid x_{n+1}=k,x_n=i)\ P(x_{n+1}=k\mid x_n=i)$$

$$=\sum_k P(x_{n+m+1}=j,x_{n+1}=k\mid x_n=i)=P(x_{n+m+1}=j,\mid x_n=i).$$

Markov **性的等价性质** 5

对状态空间S上的任意有界实值函数f有

$$E(f(x_{n+1})|x_0=i_0,\cdots,x_n=i_n)=E(f(x_{n+1})|x_n=i_n)。\qquad (5.\ 5)$$

((5. 1) **式是** (5. 5)**式的特例**, **即** $f(x)=I_{\{j\}}(x)$ **的情形**. **从** (5. 1)**式推广为**(5. 5)**式需要用测度论的基本知识**, **我们把它略去**).

[**注**] **第** 1,2,3,5 **种等价性质都有与第** 4 **种等价性质类似的表达形式,请读者自检.**

Markov **性的等价性质** 6 **(最一般的形式)**

对于常见的实数集合Λ,及由随机序列x_m在时刻n及其后的信息所决定的随机变量

h_n, **恒有**

$$P(h_n\in\Lambda|x_0=i_0,\cdots,x_n=i_n)=P(h_n\in\Lambda|x_n=i_n)$$

或

$$E(h_n|x_0=i_0,\cdots,x_n=i_n))=E(h_n|x_n=i_n)\,.$$

[**注**] **这个等价条件的内涵十分丰富, 其等价性的证明在测度论中非常典型,本书从略**.

在以实际问题为背景的 Markov **链的研究中,人们最关心的是**, **经过时间** n, **过程到达某些状态的概率有多大,以及需要多长时间才能到达这些状态**. **这类问题的描述首先涉及** Markov **链的转移特性——**Markov **链的** n **步转移概率矩阵族.**

1.2 概率转移矩阵

定义5.2 记

$$p_{ij}(n,m)=,$$

并定义无穷矩阵

$$\mathbf{P}(n,m)=(p_{ij}(n,m)).$$

由于此无穷矩阵的分量都是非负的且不超过 1，易见这种无穷矩阵的乘法满足结合律．又因为

$$p_{ij}(n,n)=d_{ij}\overset{\Delta}{=}\begin{cases}1 & (i=j)\\ 0 & (i\neq j)\end{cases},\quad \text{(Kroneker 记号)}$$

所以 $\mathbf{P}(n,n)=\mathbf{I}$ (无穷单位矩阵)．特别地，$\mathbf{P}(n,n+1)$ 称为时刻 n (到时刻 $n+1$)的概率转移矩阵 而 $\mathbf{P}(n,n+k)$ 就称为从 n 出发的 k 步概率转移矩阵．

命题５．３　概率转移矩阵族满足以下的性质

记 1 为分量全是 1 的无穷行向量(矩阵)，其维数与此 Markov 链的状态数一致．我们有

(P.1) $0\le p_{ij}(n,m)\le 1$，$\mathbf{P}(n,m)\,\mathbf{1}^{T}=\mathbf{1}^{T}$．(上标 T 表示转置运算)

(P.2)（Chapman-Kolmogorov 方程） 对于 $\forall l\ge m\ge n$ 有

$$\mathbf{P}(n,l)=\mathbf{P}(n,m)\ \mathbf{P}(m,l),\qquad (5.\ 6)$$

写成分量形式就是

$$p_{ij}(n,l)=\sum_k p_{ik}(n,m)p_{kj}(m,l).\qquad (5.\ 6)'$$

证明　验证(P. 1)

$$\sum_j p_{ij}(n,m)=\sum_j P(x_m=j\mid x_n=i).$$

$$=P(\bigcup_j\{x_m=j\}\mid x_n=i)=P(\text{全集}\Omega\mid x_n=i)=1.$$

验证(P. 2)

$$\sum_k p_{ik}(n,m)p_{kj}(m,l)=\sum_k P(x_m=k\mid x_n=i)P(x_l=j\mid x_m=k)$$

$$=\sum_k P(x_m=k\mid x_n=i)P(x_l=j\mid x_m=k,x_n=i)=\sum_k P(x_l=j,x_m=k\mid x_n=i)$$

$$=P(\bigcup_k\{x_l=j,x_m=k\}\mid x_n=i)=P(x_l=j\mid x_n=i).$$

１．３　时齐的 Markov 链

定义５．４　Markov 链称为时齐的，如果其概率转移阵 $\mathbf{P}(n,n+1)$ 与 n 无关(即 1 步转移概率与出发时刻 n 无关．

我们把此矩阵简记为 $\mathbf{P}=(p_{ij})$．由 (5. 6)式得到

$$\mathbf{P}(n,n+m)=\mathbf{P}(n,n+1)\ \mathbf{P}(n+1,n+2)\cdots\ \mathbf{P}(n+m-1,n+m)=\mathbf{P}^{m}$$

它也不依赖出发时刻 n，我们把它改记成 $\mathbf{P}^{(m)}$, ($\mathbf{P}^{(m)}=(p_{ij}^{(m)})$)，其分量

$p_{ij}^{(m)}=P(x_{n+m}=j|x_n=i)$ 为 Markov 链 $\{x_n,n\ge 0\}$ 的 m 步转移概率，它也与出发时刻 n 无

关．于是 Chapman-Kolmogorov 方程的矩阵形式变为

$$\mathbf{P}^{(m+n)} = \mathbf{P}^{(m)}\mathbf{P}^{(n)} \tag{5. 6)''}$$

而其分量形式为

$$p_{ij}^{(m+n)} = \sum_k p_{ik}^{(m)} p_{kj}^{(n)} \tag{5. 6)'''}$$

(其实 (5. 6)''式是显然的，因为它就是 $\mathrm{P}^{m+n} = \mathrm{P}^m\ \mathrm{P}^n$).

以后在本书中除非特别声明，我们所考虑的 Markov 链，均为时齐的 Markov 链．

对于时齐的 Markov 链$\{x_n, n \ge 0\}$，描述它的演化的最基本的量，就是它的转移概率矩阵 **P**=(p_{ij}),$(p_{ij}, i, j \in S)$，其中第i行就是给定$x_n = i$时，x_{n+1}的条件概率分布.

转移概率矩阵 **P** 是一个随机矩阵，即它满足：

(1) **P** 中的元素均为非负，即$p_{ij} \ge 0$，

(2) **P** 中的每一行的元素之和均为 1，即$\sum_j p_{ij} = 1$.

Markov 链的转移概率$\{p_{ij}, i, j \in S\}$（或者说转移概率矩阵 **P**）是刻画 Markov 链的统计特征的基本量. 事实上，一个 Markov 链可由其初始状态x_0的统计性质（即其初始分布$m_i^{(0)} = P(x_0 = i)$）以及其转移概率矩阵 P 所完全确定. 这就是下面的定理。

定理 5．5（Markov 链的有限维分布）若 Markov 链$\{x_n, n \ge 0\}$的初始分布为$m_i^{(0)} = P(x_0 = i)$，其转移概率矩阵为 **P**=(p_{ij})，则$\{x_n, n \ge 0\}$的有限维分布为

$$P(x_0 = i_0, x_1 = i_1, \cdots, x_n = i_n) = m_{i_0}^{(0)} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n} \qquad \forall\ i_0, i_1, \cdots, i_n \in S .$$

证明　由乘法公式与 Markov 性得

$$\begin{aligned}&P(x_0 = i_0, x_1 = i_1, \cdots, x_n = i_n)\\ &= P(x_0 = i_0)P(x_1 = i_1 | x_0 = i_0)P(x_2 = i_2 | x_0 = i_0, x_1 = i_1) \cdots P(x_n = i_n | x_0 = i_0, \cdots x_n = i_n)\\ &= m_{i_0}^{(0)} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n}\end{aligned}$$

?

记$m_i^{(n)} = P(x_n = i)$，它称为 Markov 链在时刻n的绝对概率． 再记由$m_i^{(n)}$构成的行向量为$m^{(n)} = (m_i^{(n)} : i \in S)$，那么，我们有

定理 5．6　$m^{(m+n)} = m^{(m)}\, \mathbf{P}^{(n)}$，从而有　$m^{(n)} = m^{(0)}\, \mathbf{P}^n$

证明　$m_j^{(m+n)} = P(x_{m+n} = j) = \sum_i P(x_{m+n} = j | x_m = i) P(x_m = i) = \sum_i m_i^{(m)} p_{ij}^{(n)}$

$= (m^{(m)} \mathbf{P}^{(n)})_j$.

?

定理５．５与定理５．６说明了 Markov 链的统计性质（包括其长时间极限行为），完全可由其转移概率矩阵 **P**，以及它的初始分布 $m^{(0)}$ 所决定. 因此，Markov 链的很多性质的研究就归结为随机矩阵的性质.

１．４ Markov 链的例

例５．７ （随机徘徊）

独立同分布的随机变量的部分和序列，称为随机徘徊，它是时间参数离散情形时的时齐的独立增量过程．又若其中的随机变量只取１与－１两个值，则称为简单随机徘徊.

今考虑一个简单随机徘徊$\{x_n, n \ge 0\}$，其状态空间为$S = \mathbf{Z}$ =$\{$全体整数$\}$，由x_n的定义

$$x_n = x_0 + \sum_{k=1}^{n} Z_k \text{，}$$

其中$\{Z_k, k \ge 0\}$为独立同分布随机变量序列，满足

$$Z_k \sim \begin{pmatrix} -1 & 1 \\ q & p \end{pmatrix} \quad (q = 1 - p).$$

这里x_n表示一个粒子每次分别以概率 p 与 q 向右与向左走一格. $p = \frac{1}{2}$ 的情况称为对称简单随机徘徊. 由于随机徘徊是时齐的独立增量过程, 由第 3 章可知它也是时齐的 Markov 链. 又因为$x_0, x_1, \cdots x_n$都是 $Z_0, Z_1, \cdots, Z_n$ 的部分和, 所以, 它们和 Z_{n+1} 独立，故

$$p_{ij} = P(x_{n+1} = j | x_n = i) = P(Z_{n+1} + x_n = j | x_n = i)$$

$$= P(Z_{n+1} = j - i | x_n = i) = P(Z_{n+1} = j - i) = \begin{cases} p & j = i+1 \\ q & j = i-1 \\ 0 & \text{其它} \end{cases}.$$

即其转移矩阵为

$$\mathbf{P} = (p_{ij}) = \begin{pmatrix} \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & q & 0 & p & 0 & 0 & \vdots \\ \vdots & 0 & q & 0 & p & 0 & \vdots \\ \vdots & 0 & 0 & q & 0 & p & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}.$$

而绝对概率为

$$P(x_n = j) = \begin{cases} C_n^{\frac{n+j}{2}} p^{\frac{n+j}{2}} (1-p)^{\frac{n-j}{2}}, & (\text{若}\, n \ge |j|\text{，且}\, n\, \text{与}\, j\, \text{奇偶同}) \\ 0 & \text{，(其它情形)} \end{cases}.$$

事实上 $P(x_n = j)$ 就是: n 个相互独立的随机事件$\{Z_1 = 1\}, \{Z_2 = 1\}, \ldots, \{Z_n = 1\}$中恰好有

$\frac{n+j}{2}$个发生的概率 (x_n要到达状态 j，所需的步数 n 不可能小于$|j|$，故$n \ge |j|$， 设N_n^+和 N_n^- 分别表示在前 n 步中向右和向左的步数，则显然有 $n = N_n^+ + N_n^-$ ，$x_n = N_n^+ - N_n^-$． 于是 $N_n^+ = \frac{1}{2}(n+x_n)$， 从而 $x_n = j \Leftrightarrow N_n^+ = \frac{1}{2}(n+j)$．又因为$2N_n^+ = (n+x_n)$是偶数，因此 n 与x_n 的奇偶性相同). 对于初始状态为$x_0 = i$ 的简单随机徘徊，类似可得

$$p_{ij}^{(n)} = P(x_n = j \mid x_0 = i)$$
$$= \begin{cases} C_n^{\frac{n+j-i}{2}} p^{\frac{n+j-i}{2}} (1-p)^{\frac{n-j+i}{2}} & (\text{若}n \ge |j-i|\text{，且}n\text{与}j-i\text{奇偶同}). \\ 0 & (\text{其它情形}) \end{cases}$$

例５．８ (两端为反射壁的随机徘徊) 在例５．７中, 如果在位置 a 与 b ($a<b$) 分别设置一个反射壁, 即当粒子到达 a 与 b 时, 下一步则以概率为 1 地分别 ”反射” 到 $a+1$ 与$b-1$． 此时粒子的运动仍然是一个 Markov 链, 与例５．７不同处仅在于

$$p_{a,a+1} = 1, \quad p_{a,j} = 0, \quad (j \ne a+1); \quad p_{b,b-1} = 1, \quad p_{b,j} = 0, \quad (j \ne b-1).$$

即其转移矩阵为

$$\mathrm{P} = (p_{ij}) = \begin{pmatrix} 0 & 1 & \cdots & \cdots & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & q & 0 & p & 0 & 0 & \vdots \\ \vdots & 0 & q & 0 & p & 0 & \vdots \\ \vdots & 0 & 0 & q & 0 & p & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & 1 & 0 \end{pmatrix}.$$

读者可自行写出一端反射壁的随机徘徊的转移矩阵 P.

例５．９ (两端为吸收壁的随机徘徊) 若 Markov 链取值于$\{a, a+1, \cdots, b\}$， 且其转移矩阵为

$$\mathrm{P} = (p_{ij}) = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & q & 0 & p & 0 & 0 & \vdots \\ \vdots & 0 & q & 0 & p & 0 & \vdots \\ \vdots & 0 & 0 & q & 0 & p & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & 0 & 1 \end{pmatrix},$$

则称此 Markov 链为在a与b设有吸收壁的随机徘徊．即： $p_{aa} = p_{bb} = 1$，并称a与b为

Markov 链的吸收态.

例5．10 （货仓存货问题）

设一个运货仓库每月进货的件数是一个独立同分布的随机变量序列$\{h_n : n=1,2,\cdots\}$，其中h_n表示第 n 个月进货的件数，$h_n \sim \begin{pmatrix} 1 & \cdots & N \\ p_1 & \cdots & p_N \end{pmatrix}$. 又设$x_0 \sim \begin{pmatrix} 1 & \cdots & N \\ \frac{1}{N} & \cdots & \frac{1}{N} \end{pmatrix}$，仓库的货物容量为 N 件，每当仓库的货物超过N件时，就将N件打包发运. 记第n个月底的存货量为x_n，那么

$$x_n = \begin{cases} x_{n-1} + h_n & h_n < N - x_{n-1} \\ x_{n-1} + h_n - N & h_n \ge N - x_{n-1} \end{cases}。$$

于是$\{x_n:: n \ge 0\}$是一个 Markov 链， 事实上

$$\begin{aligned}
&P(x_{n+1} = j | x_0 = i_0, x_1 = i_1, \cdots, x_{n-1} = i_{n-1}, x_n = i) \\
&= \begin{cases} P(i + h_{n+1} = j \mid x_0 = i_0, x_1 = i_1, \cdots, x_{n-1} = i_{n-1}, x_n = i) & (i < j < N) \\ P(i + h_{n+1} - N = j \mid x_0 = i_0, \cdots, x_{n-1} = i_{n-1}, x_n = i) & (i \ge j) \end{cases}. \\
&= \begin{cases} P(h_{n+1} = j - i \mid x_0 = i_0, \cdots, x_{n-1} = i_{n-1}, x_n = i) & (i < j < N) \\ P(h_{n+1} = N + j - i \mid x_0 = i_0, \cdots, x_{n-1} = i_{n-1}, x_n = i) & (\quad j \le i) \end{cases}
\end{aligned}$$

$$= \begin{cases} P(h_{n+1} = j - i) & (i < j < N) \\ P(h_{n+1} = N + j - i) & (i \ge j) \end{cases} = \begin{cases} p_{j-i} & (i < j < N) \\ p_{N+j-i} & (i \ge j) \end{cases}.$$

同样的推理也得到

$$P(x_1 = j \mid x_0 = i) = \begin{cases} p_{j-i} & (i < j < N) \\ p_{N+j-i} & (i \ge j) \end{cases}.$$

所以$\{x_n, n \ge 0\}$是时齐的 Markov 链，且其转移矩阵的分量为

$$p_{ij} = \begin{cases} p_{j-i} & (i < j < N) \\ p_{N+j-i} & (i \ge j) \end{cases}.$$

例5．11 （品牌选择） 设市场上销售 A， B，C，D 四种品牌的牙膏，根据市场调查，消费者购买哪一种品牌的牙膏，近似地仅与他前一次购买的品牌有关，而与这之前购买的品牌无关．记x_0为某消费者最初所购买的牙膏的品牌，$x_1, x_2, \cdots$分别表示他在这之后各次所购买的牙膏的品牌，则$\{x_n : n \ge 0\}$为一 Markov 链，其状态空间为 S={A,B,C,D}，它的转移概率矩阵可以从市场调查中近似地获得.假定经过长期市场调查统计得到近似的转移矩阵为

$$\mathbf{P}=\begin{pmatrix} 0.90 & 0.05 & 0.03 & 0.02 \\ 0.10 & 0.80 & 0.05 & 0.05 \\ 0.08 & 0.10 & 0.80 & 0.02 \\ 0.10 & 0.10 & 0.10 & 0.70 \end{pmatrix}.$$

对于这个链，人们感兴趣的是这四种品牌的牙膏的市场占有率随时间进程的变化.

例５．１２ （Wright-Fisher 遗传模型 ） 遗传的要素是染色体．遗传性质的携带者称为基因，它们位于染色体上. 基因控制着生物的特征，它们是成对出现的. 控制同一特征的不同基因称为等位基因，记这对等位基因为 A 和 a.，分别称为显性的与隐性的．在一个总体中基因 A 和 a 出现的频率称为基因频率，分别记为 p 和 $1-p$ ． 设总体中的个体数为 $2N$. 每个个体的基因按 A 型基因的基因频率的大小, 在下一代中转移成为 A 型基因. 因此，繁殖出的第二代的基因型是由试验次数为 $2N$ 的 Bernoulli 试验所确定的. 即如果在第 n 代母体中 A 型基因出现了 i 次，而 a 型基因出现了 $2N-i$ 次, 则下一代出现 A 型基因的概率为 $p_i=\dfrac{i}{2N}$ ，而出现 a 型基因的概率为 $1-p_i$. 记 x_n 为第 n 代中携带 A 型基因的个体数，则易证 $\{x_n,n\geq 0\}$ 是一状态空间为 $S=\{0,1,2,\cdots,2N\}$ 的时齐 Markov 链，其转移概率矩阵为 $\mathbf{P}=(p_{ij})$，其中

$$p_{ij}=P(x_{n+1}=j|x_n=i)=C_{2N}{}^{j}p_i{}^{j}(1-p_i)^{2N-j}=C_{2N}{}^{j}(\frac{i}{2N})^{j}(1-\frac{i}{2N})^{2N-j}.$$

2．Markov 链的状态分类

2. 1 首达分解, n 步转移概率的递推式，矩母函数，常返性

定义５.１３（首次进入状态 j 的时刻） 把从 i 出发在 $n(\geq 1)$ 步转移中首次到达 j 的概率记为 $f_{ij}^{(n)}$．用数学式表达即

$$f_{ij}^{(n)}=P(x_n=j,x_k\neq j,k=1,2,\cdots n-1|x_0=i),(n\geq 1) \qquad (5.\ 7)$$

而把 Markov 链 $\{x_n,n\geq 0\}$ 首次击中状态 j 的时刻记为 T_j，那么 T_j 是一随机变量, 但是与普通的随机变量有一点不同, 就是它可以取值 ∞ （事实上，它还是一个 (x_n) 停时，即它是否不大于 m ，只由随机序列 x_n 在时刻 m 前的信息完全确定)，即

$$T_j=\begin{cases}\min\{n\geq 1: x_n=j\} & (\text{若存在} n\geq 1 \text{使得} x_n=j) \\ \infty & (\text{其它情形})\end{cases}.$$

从而有

$$P(T_j=n|x_0=i)=f_{ij}^{(n)} \qquad (n\geq 1).$$

$$P(T_j = \infty | x_0 = i) = 1 - \sum_{n=1}^{\infty} f_{ij}^{(n)} \quad . . \tag{5.8}$$

再记

$$f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)} = P(T_j < +\infty | x_0 = i) = \mathrm{P}(\exists n \geq 1 \text{使得} x_n = i | x_0 = i) \leq 1, \tag{5.9}$$

那么 f_{ij} 表示 $\{x_n : n \geq 0\}$ 从 i 出发，在有限时间内它能够到达 j 的概率（或者说，从 i 出发最终转移到状态 j 的概率）.

定义5.14（常返态与暂态） Markov 链 $\{x_n : n \geq 0\}$ 的状态 i 称为常返态，如果从 i 出发，能以概率为 1 地，最终又能回到 i，即 $f_{ii} = 1$. 如果状态 i 不是常返态，则称它为暂态.

定理5.15（Markov 链的首达分解定理） 对于任意状态 i, j，任意时刻 n，有

$$p_{ij}^{(n)} = \sum_{k=1}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)} . \tag{5.10}$$

证明 利用 Markov 性质，我们有

$$p_{ij}^{(n)} = P(x_n = j | x_0 = i) = \sum_{k=1}^{n} P(x_n = j, T_j = k | x_0 = i))$$

$$= \sum_{k=1}^{n} P(x_n = j | T_j = k, x_0 = i) P(T_j = k | x_0 = i) = \sum_{k=1}^{n} P(x_n = j | x_k = j) f_{ij}^{(k)} . ?$$

令

$$p_{ij}^{(0)} = d_{ij}, f_{ij}^{(0)} = 0,$$

并定义矩母函数

$$P_{ij}(z) = \sum_{k=0}^{\infty} p_{ij}^{(n)} z^n, \quad F_{ij}(z) = \sum_{k=0}^{\infty} f_{ij}^{(n)} z^n .$$

易见定理5．15可以写成如下的矩母函数形式

定理5.15'

$$P_{ij}(z) = d_{ij} + F_{ij}(z) P_{jj}(z) . \tag{5.10'}$$

. ?

从（5．10）' 可以解出

$$P_{jj}(z) = \frac{1}{1 - F_{jj}(z)}, \tag{5.11}$$

$$F_{ij}(z) = \frac{F_{ij}(z)}{1 - F_{jj}(z)} . \tag{5.12}$$

在(5．11)式中令$z \to 1-0$，便得到(利用数学分析中的 Abel 定理)

$$\sum_{n=1}^{\infty} p_{jj}^{(n)} = \frac{1}{1-f_{jj}}. \qquad (5.11)'$$

于是，由常返性的定义立得如下的充要条件

定理 5.16 （常返性与暂态的条件）

$$j \text{ 为常返态} \iff \sum_{n=1}^{\infty} p_{jj}^{(n)} = \infty;$$

$$j \text{ 为暂态} \iff \sum_{n=1}^{\infty} p_{jj}^{(n)} < \infty.$$

推论5．17 状态j为暂态 $\Rightarrow \forall i, \ \sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty$，从而 $\lim_{n\to\infty} p_{ij}^{(n)} = 0$.

证明 右$= F_{ij}(1) < \infty$.

推论5．18 (有限状态 Markov 链) 有限状态 Markov 链至少存在一个常返态．

证明 假设有限状态 Markov 链的所有状态均为暂态, 那么由

$$1 = \sum_{1\le j\le N} P(x_n = j \mid x_0 = i) = \sum_{1\le j\le N} p_{ij}^{(n)} \xrightarrow{n\to\infty} 0 .$$

这就导致了矛盾．

定义5.20 如果存在$n \ge 0$，使$p_{ij}^{(n)} > 0$，则称状态i可达状态j，记作$i \to j$， 表示从状态i经过有限步的转移可以到达状态j. 它等价于：存在$i_1,\cdots,i_{n-1}$使$p_{i_k i_{k+1}} > 0, 0 \le k \le n-1$, 其中$i_0 = i, i_n = j$.

注 （Markov 链转移的图示） 把 Markov 链的状态记为顶点。如果$p_{ij} > 0$，则从状态i到状态j画一条有向边．这样就得到表示 Markov 链的转移情况的一个有向图（有些边可能是双向的）．那么，$i \to j$ 等价于存在顶点i到顶点j的一条由定向边组成的通路．

命题 5.21 设 $i \to j$ 且 $j \to i$ （记为$i \leftrightarrow j$），则状态i与j常返与否是相同的.

证明 由于$i \leftrightarrow j$，故存在$m, n \ge 0$，使得$p_{ij}^{(m)} > 0, \quad p_{ji}^{(n)} > 0$，记

$$a \overset{\Delta}{=} p_{ij}^{(m)} p_{ji}^{(n)} > 0 .$$

由 Chapman-Kolmogorov 方程可知，对任意非负整数r，有

$$p_{ii}^{(m+r+n)} \ge p_{ij}^{(m)} p_{jj}^{(r)} p_{ji}^{(n)} = a \cdot p_{jj}^{(r)} .$$

由对称性同样有

$$p_{jj}^{(n+r+m)} \ge p_{ji}^{(n)} p_{ii}^{(r)} p_{ij}^{(m)} = a \cdot p_{ii}^{(r)} .$$

从而可知级数$\sum_r p_{ii}^{(r)}$ 与$\sum_r p_{jj}^{(r)}$ 同时收敛或发散. 故i, j 的常返与否是相同的

例５．２２ (二值链)

$$\mathrm{P}=\begin{pmatrix} a & 1-a \\ 1-b & b \end{pmatrix} \qquad (0<a,b<1) .$$

直观上可以想到状态 0 和 1 均是常返态. 下面证明此结论.

利用归纳法我们可以得到

$$\mathbf{P}^{(n)}=\mathbf{P}^n=\frac{1}{2-(a+b)}\begin{pmatrix}1-b & 1-a \\ 1-b & 1-a\end{pmatrix}+\frac{(a+b-1)^n}{2-(a+b)}\begin{pmatrix}1-a & -(1-a) \\ -(1-b) & 1-b\end{pmatrix}.$$

故 $$p_{00}^{(n)}=\frac{1-b}{2-(a+b)}+\frac{1-a}{2-(a+b)}(a+b-1)^n .$$

由于$0<a,b<1$，就有 $\lim\limits_{n\to\infty} p_{00}^{(n)}=\dfrac{1-b}{2-(a+b)}>0$. 从而 $\sum\limits_{n=1}^{\infty} p_{00}^{(n)}=\infty$, 故 0 为常返态. 同样状态 1 也是常返态.

例５．２３ 在简单随机徘徊中考虑状态 0. 此时

$$p_{00}^{(2n+1)}=0 \qquad (n=0,1,2,\cdots) , \quad p_{00}^{(2n)}=C_{2n}^n p^n q^n \qquad (n=1,2,\cdots) .$$

用 Stirling 公式：$n!\sim\sqrt{2p}\, n^{n+\frac{1}{2}}e^{-n}$ （其中当$\lim\limits_{n\to\infty}\dfrac{a_n}{b_n}=1$时，我们记$a_n\sim b_n$），我们可得

$$p_{00}^{(2n)}\sim\frac{(4pq)^n}{\sqrt{pn}} .$$

因此$\sum\limits_{n=1}^{\infty} p_{00}^{(n)}$ 收敛当且仅当$\sum\limits_{n=1}^{\infty}\dfrac{(4pq)^n}{\sqrt{pn}}$ 收敛．又因为$4pq\le 1$，且等号成立当且仅当

$p=q=\dfrac{1}{2}$，于是$\sum\limits_{n=1}^{\infty} p_{00}^{(n)}=\infty$ 当且仅当 $p=q=\dfrac{1}{2}$．也就是说，当且只当在简单随机徘徊为对称时，状态 0 是常返的．

*__例５．２４__（Z^d 上的对称随机徘徊） 记Z^d为 *d* 维整数格点组成的集合．每一个格点有 *d* 个方向，故各有 *2d* 个邻点．在 Markov 链已知处在某一个格点的条件下，以相同的条件概率(即各以概率 $\dfrac{1}{2d}$)，在下一个时刻转移到它的任意一个邻点，这样的 Markov 链称为 Z^d 上的对称随机徘徊．显见各个点的常返性与暂态性是一样的．下面我们考察原点的情形．显然 $p_{00}^{(2n-1)}=0$.

对于正整数m，把m 个不同的元素分成 l 个组，使第 j 组恰有k_j 个元素($k_1+\cdots+k_l=m$)的不

同分法的数目记为$C_m^{k_1,\cdots,k_l}$． 那么，由归纳法可得$C_m^{k_1,\cdots,k_l} = C_m^{k_1} C_{m-k_1}^{k_2} \cdots = \dfrac{m!}{k_1!\cdots k_l!}$．于是

$$p_{00}^{(n)} = \sum_{k_1+\cdots+k_d=n} P(\bigcap_{i=1}^{d}\{\text{向第}i\text{个坐标方向两侧独立地各走}k_i\text{步}\})$$

$$= \sum_{k_1+\cdots+k_d=n} C_{2n}^{k_1,k_1,\cdots,k_d,k_d} (\frac{1}{2d})^{2n} = \sum_{k_1+\cdots+k_d=n} \frac{(2n)!}{(k_1!)^2\cdots(k_d!)^2} (\frac{1}{2d})^{2n}$$

$$= \frac{1}{2^{2n}} C_{2n}^{n} \sum_{k_1+\cdots+k_d=n} [C_n^{k_1,\cdots,k_d} \frac{1}{d^n}]^2 .$$

这样，当$d=2$时，用$\mathring{a}_{k£n} [C_n^{k,n-k}]^2 = C_{2n}^n$及Stirling公式，我们有

$$p_{00}^{(n)} = \frac{1}{4^{2n}} (C_{2n}^{n})^2 \approx \frac{1}{p \cdot n} .$$

而当$d \ge 3$时我们利用下面的引理．

引理５．２５ 在$n \to \infty$时有近似关系

$$\max_{\sum_{i=1}^{d} k_i = n} C_n^{k_1,\cdots,k_d} \approx C_n^{\frac{n}{d},\cdots,\frac{n}{d}} ,\quad \frac{1}{2^{2n}} C_{2n}^{n} \approx \frac{1}{\sqrt{p \cdot n}} . \qquad ?$$

(此引理可用Stirling公式证明)．

利用引理５．２５， 就得到

$$p_{00}^{(n)} \le \frac{1}{2^{2n}} \frac{1}{d^n} C_{2n}^{n} [\max_{\sum_{i=1}^{d} k_i = n} C_n^{k_1,\cdots,k_d}] \approx \frac{1}{d^n} \frac{1}{\sqrt{pn}} C_n^{\frac{n}{d},\cdots,\frac{n}{d}}$$

$$\approx \frac{1}{d^n \sqrt{pn}} \frac{n^n e^{-n} \sqrt{2pn}}{[(\frac{n}{d})^{\frac{n}{d}} e^{-\frac{n}{d}} \sqrt{2p\frac{n}{d}}]^d} \approx \text{常数} \frac{1}{n^{\frac{d}{2}}} .$$

可见

$$\sum_{n=0}^{\infty} p_{00}^{(n)} \begin{cases} = \infty & (d \le 2) \\ < \infty & (d \ge 3) \end{cases} .$$

即：对于Z^d上的对称简单随机徘徊，当 $d \le 2$ 时，一切状态都是常返态．而当 $d \ge 3$时，一切状态都是暂态．这个结论在直观上是很清楚的，在d大时， 能去的地方多了，回返的可能性就小了．

因为常返态能以概率为 1 地返回，直观地看，从常返态i能到达的状态 j ,必须以概率为 1 返回i，而返回i后又可达 j，看来似乎 j 总能返回自己，即 j 也应是常返态．另一方面，常返态既能以概率为 1 地返回一次，就能返回第二次,…,从而能返回∞次．即若状态i为常返态，则 Markov 链无穷次返回状态i的概率为 1．而若状态j为暂态，则 Markov 链以正概率不能返回，从而无穷次返回状态j的概率为 0．下段将致力于证明这两个结论．

2．2 常返性再访与 Markov 链的基本结构

设随机变量h_j表示 Markov 链x_n访问状态 j 的次数，即把x_n作为下标$n \ge 1$的随机序

列时，在此序列中总共取到状态 j 的次数（若在此序列中 j 被取到了 ∞ 次，则我们认为 $h_j=\infty$．所以 h_j 是一个可以取值 ∞ 的非负整值随机变量）．按定义就有

$$f_{ij}=P(h_j\ge 1|x_0=i)\quad .$$

定理 5．2 6

(1) 若 j 为常返态，则

$$P(h_j=\infty|x_0=i)=f_{ij}(\forall i\in S).$$

特别是在 $i=j$ 时，有 $P(h_j=\infty|x_0=j)=1$ ．

(2) 若 j 为暂态，则 $E(h_j|x_0=i)=\dfrac{f_{ij}}{1-f_{jj}}\quad(\forall i\in S)$．

证明 （1）按 T_j 取值的不同，把要考虑的事件划分为互不相容的事件之并，利用 Markov 性及全概率公式，我们得到

$$P(h_j\ge n|x_0=i)=\sum_{k=1}^{\infty}P(h_j\ge n|T_j=k,x_0=i)P(T_j=k|x_0=i)$$

$$=\sum_{k=1}^{\infty}P(h_j\ge n|x_k=j,x_l\ne j,(l<k),x_0=i)P(T_j=k|x_0=i)$$

$$=\sum_{k=1}^{\infty}P(h_j\ge n|x_k=j)P(T_j=k|x_0=i)=\sum_{k=1}^{\infty}P(h_j\ge n-1|x_0=j)P(T_j=k|x_0=i)$$

$$=P(h_j\ge n-1|x_0=j)\sum_{k=1}^{\infty}P(T_j=k|x_0=i)=f_{ij}P(h_j\ge n-1|x_0=j)$$

反复使用上式，就可推出

$$P(h_j\ge n|x_0=i)=f_{ij}[f_{jj}]^{n-2}P(h_j\ge 1|x_0=j)=f_{ij}[f_{jj}]^{n-1}\qquad(n\ge 1).$$

由于 j 为常返态，即 $f_{jj}=1$，令 $n\to\infty$ 就得到

$$P(h_j=\infty|x_0=i)=f_{ij}\qquad\forall i\in S$$

（2）由于 h_j 为非负整值随机变量，与普通的期望类似地有条件期望公式

$$E(h_j|x_0=i)=\sum_{n=1}^{\infty}P(h_j\ge n|x_0=i).$$

故由（1）中的推导及 $f_{jj}<1$(因为假定暂态 j 为暂态) 可推出

$$E(h_j|x_0=i)=\sum_{n=1}^{\infty} f_{ij}[f_{jj}]^{n-1}=\frac{f_{ij}}{1-f_{jj}} \qquad \forall i\in S.$$

推论５．２７ (**常返与暂态的另一个等价条件**)

(1) **状态**j**为常返态，当且仅当，**$P(h_j=\infty|x_0=j)=1$；

(2) **状态**j**为暂态，当且仅当，**$P(h_j=\infty|x_0=j)=0$.

证(2) **由定理**５．２６**得**

$P(h_j=\infty|x_0=j)=0 \quad \Leftrightarrow E(h_j|x_0=j)<\infty \quad \Leftrightarrow f_{jj}\neq 1$，**即**$j$**为暂态.**

[**注** 1] **这个推论就是说，在条件概率** $P(\cdot|x_0=j)$**下，事件** $h_j=\infty$**的条件概率只能为** 0 **或** 1， **也称为对于此事件"条件 0-1 律"成立.**

[**注** 2] **推论５．２７清楚地说明了常返态与暂态的含义．对于** Markov **链，常返态是由它出发能**∞**次返回的状态，而随着时间的发展，暂态将逐渐消失．所以，在对用** Markov **链描述的模型作稳态设计时，暂态是不予考虑的．这正说明了为什么区分常返态与暂态是十分重要的．**

定理５．２８ **若**i**常返，且**$i\to j$，**则**$f_{ji}=f_{jj}=1$，**从而**j**也常返.**

证明 **记** $m_0=\inf_m p_{ij}^{(m)}>0$．**由于**$i$**为常返态，故**

$$1=P(h_i=\infty|x_0=i)$$

$$=P(x_{m_0}=j,\text{有}\infty\text{个}n\text{使}x_{m_0+n}=i|x_0=i)+P(x_{m_0}\neq j,\text{有}\infty\text{个}n\text{使}x_{m_0+n}=i|x_0=i)$$

$$\leq P(x_{m_0}=j,\text{有}\infty\text{个}n\text{使}x_{m_0+n}=i|x_0=i)+P(x_{m_0}\neq j|x_0=i)$$

$$\leq p_{ij}^{(m_0)}P(T_i<\infty|x_0=j)+(1-p_{ij}^{(m_0)})$$

$$=1-p_{ij}^{(m_0)}(1-P\ (T_i<\infty|x_0=j))\leq 1.$$

因此

$$f_{ji}=P(T_i<\infty|x_0=j)=1.$$

于是存在n**使**$f_{ji}^{(n)}>0$，**从而** $p_{ji}^{(n)}\geq f_{ji}^{(n)}>0$，**即** $j\to i$．**由命题**５．２１**得**j**的常返性.**

上面得到的关于状态为常返（或暂态）的条件，从概念的含义看是很清楚的．但是很难用于实际判断．比较起来，用命题５．１６来判断状态j**是否常返常更为可行.**

[**注**] ($\sum_{n=1}^{\infty}p_{jj}^{(n)}$ **的概率含义**) **记**

$$I_{\{j\}}(x_n)=\begin{cases}1 & (\text{若}x_n=j)\\ 0 & (\text{其它情形})\end{cases}.$$

那么 $h_j=\sum_{n=1}^{\infty}I_{\{j\}}(x_n)$. 因此

$$\sum_{n=1}^{\infty}p_{jj}^{(n)}=\sum_{n=1}^{\infty}P(x_n=j|x_0=j)=\sum_{n=1}^{\infty}E(I_{\{j\}}(x_n)|x_0=i)=E(h_j|x_0=i).$$

Markov 链的基本结构

定义５．２８（状态的互通性） 两个互相可达的i,j称为是互通的，记为$i\leftrightarrow j$.

命题５．２９ 在常返态间，互通性是等价关系，即它满足

（１） 自反性：$i\leftrightarrow i$；

（２）对称性：若$i\leftrightarrow j$，则$i\leftrightarrow j$；

（３）传递性：若$i\leftrightarrow j$，且$j\leftrightarrow k$，则$i\leftrightarrow k$.

证明 我们只须证明传递性. 设$i\leftrightarrow j$，则存在整数 n，使得 $p_{ij}^{(n)}>0$，又由于 $j\leftrightarrow k$，故存在整数 m，使得 $p_{jk}^{(m)}>0$. 因此 由 Chapman-Kolmogorov 方程得到

$$p_{ik}^{(m+n)}=\sum_{r\in S}p_{ir}^{(n)}p_{rk}^{(m)}\geq p_{ij}^{(n)}p_{jk}^{(m)}>0.$$

故$i\to k$. 同理可证，$k\to i$，从而有$i\leftrightarrow k$.

注 与多数课本不同的是，在本书中我们只对常返态考虑等价性，原因在于对于暂态果i不可能有$i\to i$. 也就是说，只在常返态间，互通性才是自返的，即 $i\to i$. 只有此时它才可能是一个数学上的等价关系.

Markov 链的状态，除暂态外，可利用互通这个等价关系划分成不同的等价类，即把互通的常返状态归入同一类中，称为一个常返类．而把所有的暂态状态，通通放到一个“另类”中， 不再加以细分. 单个状态k是一个常返类的充要要条件为，$p_{kk}=1$，此时k为吸收态.

定义５．３０ (状态闭集与不可约性)

状态的集合A称为闭集，如果对于任意状态$i\in A$，满足 $\sum_{j\in A}p_{ij}=1$. 显见状态空间 (即全体状态) 是一个闭集. 如果除此以外再也没有其它非空闭集，则称此 Markov 链为不可约的.

显见，一个常返类是闭集，而且它不含更小的闭集.

简单随机徘徊的所有状态都在同一等价类，故它是一个不可约 Markov 链. 而两端为 0 和N的吸收壁随机徘徊，吸收状态 0 和N分别都是一个单点常返类，它们是仅有的真闭集(即除状态空间以外的闭集)，此 Markov 链是可约的.

命题５．３１（状态分解定理） （１）Markov 链的状态空间 S 可唯一分解为：

$$S=T\cup H_1\cup H_2\cup\cdots,$$

其中 T 为暂态的全体，而H_i为等价常返类.

（２）若 Markov 链的初分布集中在某个常返类H_k上，则此 Markov 链概率为 1 地永远在此常返类中，也就是说，它也可以看成状态空间为H_k的不可约 Markov 链.

所有的状态全是常返态的链称为常返链；没有常返态的链称为暂态链．简单随机徘徊当$p=q=\frac{1}{2}$时（即对称的简单随机徘徊），是一个常返链；而当$p\neq q$时，是暂态链．Z^d上的对称的简单随机徘徊当$d\leq 2$时，是一个常返链；而当$d\geq 3$时，是暂态链．

１．３　平均回访时间与正常返性

我们将用平均回访时间是否有限，来对常返态作进一步的区分．

定义５．３２（平均回访时间） 对于 Markov 链首达i的时刻T_i，记

$$m_i = E(T_i | x_0 = i),$$

则m_i表示状态i的平均回访时间．T_i是一个可取∞的广义整值随机变量，它的分布表为

$$\begin{pmatrix} 1 & 2 & \cdots & \infty \\ f_{ii}^{(1)} & f_{ii}^{(2)} & \cdots & 1-f_{ii} \end{pmatrix}.$$

当i为暂态时，因为$f_{ii}<1$，所以$m_i=\infty$；而在i为常返时则有

$$m_i = \sum_{i=1}^{\infty} nP(T_i = n | x_0 = i) = \sum_{n=1}^{\infty} n f_{ii}^{(n)}.$$

于是利用m_i我们可以对常返态作进一步的区分如下

定义５．３３ 常返态i称为零常返的（简称零态），如果$m_i=\infty$；否则称为正常返的．

（从后面的定理５.４７，可以知道有限状态 Markov 链的任何常返态必定是正常返的.可见，零常返只能出现在可数状态 Markov 链的情形）．

我们还要引入一种特殊的，而且在物理中占重要地位的一种正常返态，称为遍历态。为此我们先介绍周期的概念．

定义５．３４（状态的周期） 将满足$p_{ii}^{(n)}>0$的所有$n\geq 1$的整数的最大公约数记成d_i．(如果对所有$n\geq 1$都有$p_{ii}^{(n)}=0$，则约定$d_i=\infty$)． $d_i=1$或∞的状态，称为非周期的，$d_i>1$ 的状态称为周期的．

由周期的定义立即可知，若 n 不能被d_i整除，则必有$p_{ii}^{(n)}=0$。

定义５．３５ 正常返的非周期状态称为遍历态．

例５．２２（续） 例５．２２中的两个状态均为正常返的，又由于$p_{ii}^{(1)}>0$（i=0,1），即$d_i=1$（i=0,1），故此 Markov 链的状态 0 和 1 均为遍历态．

读者可自行验证下面的定理．

定理 5．３６ 设$i\leftrightarrow j$，则

（1） i 为常返态，当且仅当 j 为常返态．
（2） i 为暂态，当且仅当 j 为暂态．
（3） i 为零常返态，当且仅当 j 为零常返态．
（4） i 为正常返态，当且仅当 j 为正常返态．
（5） i 为遍历态，当且仅当 j 为遍历态．

[注] 事实上，此时我们还有：状态 i 为周期态，当且仅当 j 为周期态，且 $d_i = d_j$。其证明方法是用初等数论，本书略去.

以上结论说明，零常返，正常返，周期大小，遍历，都是常返等价类的“类性质”，它们对同一个等价类中的状态都是相同的．

3． Markov 链的转移概率的极限与不变分布

3.1 不变分布与平稳 Markov 链

定义５．３７ 状态空间 S 上的概率分布 $p = \{p_1, p_2, \cdots\}$ 称为 P 的不变概率分布，简称不变分布，如果 $p = p\mathbf{P}$.

不变分布未必存在．若状态空间 S 为有限集，且不变分布存在，则 1 是矩阵 P 的左特征值，而不变分布 p 是它的左特征向量．此时 p 可由代数方程组

$$p = p\mathbf{P}, \quad p\ \mathbf{1}^{\mathrm{T}} = 1$$

的非负解得到，其中 $\mathbf{1} = (1, \cdots, 1)$（上标的 T 表示取转置）.

命题５．３８ 若转移矩阵为 $\mathbf{P}$ 的 Markov 链 $\{x_{\mathrm{n}}, \mathrm{n} \ge 0\}$ 的初始分布取为不变分布 p，则 $\{x_{\mathrm{n}}, \mathrm{n} \ge 0\}$ 为平稳列，即对于 $\forall \mathrm{n}, \mathrm{k}, \mathrm{n}_1, \cdots, \mathrm{n}_{\mathrm{k}}$，$(x_{\mathrm{n}_1+\mathrm{n}}, \cdots, x_{\mathrm{n}_{\mathrm{k}}+\mathrm{n}}\}$ 与 $(x_{\mathrm{n}_1}, \cdots, x_{\mathrm{n}_{\mathrm{k}}}\}$ 有相同的联合分布.（关于平稳列的一般理论可参见第 11 章).

(请自行验证).

3.2 有限状态 Markov 链的的不变分布与极限分布

定理５．３９ 若 Markov 链 $\{x_n, n \ge 0\}$ 的状态空间 S 为有限集（不妨设 $S = \{1, 2, \cdots, N\}$），且其转移概率矩阵 P=(p_{ij}) 满足 $p_{ij} > 0 \quad \forall i, j \in S$，则存在 S 上唯一的概率分布 $p = \{p_1, p_2, \cdots, p_N\}$，使得对所有 $i, j \in S$ 及 $d = \min(p_{ij} : i, j \in S) \le \frac{1}{N}$，都有：

（1） $p = p\mathbf{P}$ （即 p 为 P 的不变概率分布），而且 $p_j \ge d$；

（2） $\lim_{n\to\infty} \mathbf{P}^n = p\ \mathbf{1}^{\mathrm{T}}$，

此外，P 还满足指数遍历性，即

$$\left| p_{ij}^{(n)} - p_j \right| \le (1 - Nd)^n .$$

[注] (2) 远比 (1) 强.

证明 记 $M_j^{(n)}$ 与 $m_j^{(n)}$ 分别为矩阵 $\mathbf{P}^n$ 的第 j 列的最大与最小值，显然

$$d \equiv \min(p_{ij} : i, j \in S) \le p_{ij} = 1 - \sum_{k \ne j} p_{ik} \le 1 - (N-1)d \;.$$

因此

$$d \le \frac{1}{N}, \quad m_j^{(1)} \ge d \quad , \quad M_j^{(1)} \le 1 - (N-1)d \,.$$

进而

$$M_j^{(1)} - m_j^{(1)} \le 1 - (N-1)d - d = 1 - Nd \;. \qquad (\text{E.1})$$

再则，我们有

$$M_j^{(n+1)} = \max_i p_{ij}^{(n+1)} = \max_i \sum_k p_{ik} p_{kj}^{(n)} \le \max_i \sum_k p_{ik} M_j^{(n)} = M_j^{(n)} \,.$$

类似地

$$m_j^{(n+1)} = \min_i p_{ij}^{(n+1)} = \min_i \sum_k p_{ik} p_{kj}^{(n)} \ge \min_i \sum_k p_{ik} m_j^{(n)} = m_j^{(n)}$$

这说明 $M_j^{(n)}$ 是单调下降的，而 $m_j^{(n)}$ 是单调上升的. 我们将证明

$$0 \le M_j^{(n)} - m_j^{(n)} \le (1 - Nd)^n \to 0 \quad (n \to \infty) \,. \qquad (\text{E.2})$$

只要证明了（E.2）， 就由它可以推出，当 $n \to \infty$ 时，$\mathbf{P}^n$ 的第 j 列的各分量与 $M_j^{(n)}$ 及 $m_j^{(n)}$ 有相同的极限 （令 $p_j \equiv \lim_{n\to\infty} M_j^{(n)}$，则由 $m_j^{(n)} - M_j^{(n)} \le p_{ij}^{(n)} - p_j \le M_j^{(n)} - m_j^{(n)}$， 直接导出 $\lim_{n\to\infty} p_j^{(n)} = p_j$）. 从而有

$$p_j = \lim_{n\to\infty} p_{ij}^{(n+1)} = \lim_{n\to\infty} \sum_{k\in S} p_{ik}^{(n)} p_{kj} = \sum_{k\in S} p_k p_{kj}$$

可见为了证明(1)，只须证明（E.2）. 现在我们证明（E.2）. 注意

$$p_{ij}^{(n+1)} - p_{i^*j}^{(n+1)} = \sum_k p_{ik} p_{kj}^{(n)} - \sum_k p_{i^*k} p_{kj}^{(n)} = \sum_k (p_{ik} - p_{i^*k}) p_{kj}^{(n)}$$

$$= \sum_{k\in J} (p_{ik} - p_{i^*k}) p_{kj}^{(n)} + \sum_{k\in J^*} (p_{ik} - p_{i^*k}) p_{kj}^{(n)} \,,$$

其中 $J = \{ j \in S : p_{ij} > p_{i^*j} \}$， $J^* = \{ j \in S : p_{ij} \le p_{i^*j} \}$. 由于

$$\sum_{j\in J} (p_{ij} - p_{i^*j}) + \sum_{j\in J^*} (p_{ij} - p_{i^*j}) = \sum_{j\in S} (p_{ij} - p_{i^*j}) = 1 - 1 = 0 \,,$$

故而

$$\sum_{j\in J^*} (p_{ij} - p_{i^*j}) = -\sum_{j\in J} (p_{ij} - p_{i^*j}) \,,$$

并且

$$\sum_{j\in J} (p_{ij} - p_{i^*j}) = \sum_{j\in J} p_{ij} - \sum_{j\in J} p_{i^*j} = 1 - \sum_{j\in J^*} p_{ij} - \sum_{j\in J} p_{i^*j} \le 1 - Nd \;.$$

从而

$$p_{ij}^{(n+1)} - p_{i^*j}^{(n+1)} \le \sum_{k\in J}(p_{ik} - p_{i^*k})M_{kj}^{(n)} + \sum_{k\in J^*}(p_{ik} - p_{i^*k})m_{kj}^{(n)}$$

$$= (M_j^{(n)} - m_j^{(n)})\sum_{k\in J}(p_{ik} - p_{i^*k}) \le (1 - Nd)(M_j^{(n)} - m_j^{(n)}) . \qquad \text{(E.3)}$$

若选取 i, i^* 满足 $p_{ij}^{(n+1)} = M_j^{(n+1)}$， $p_{i^*j}^{(n+1)} = m_j^{(n+1)}$，则由（E.3）得

$$M_j^{(n+1)} - m_j^{(n+1)} \le (1 - Nd)(M_j^{(n)} - m_j^{(n)}) . \qquad \text{(E.4)}$$

这样迭代多次，再利用（E.4）便得（E.2）.

最后我们证明不变分布是唯一的．设 $m = \{m_{1,}, m_2, \cdots, m_n\}$ 是 $\mathbf{P}$ 的另一个不变分布，则有

$$m = m\mathbf{P} = m\mathbf{P}^2 = \cdots m\mathbf{P}^{\mathbf{n}} ® m(\mathbf{1}^{\mathrm{T}}p) = (m\mathbf{1}^{\mathrm{T}})p = p .$$

例５．２２（续）

$$\mathrm{P} = \begin{pmatrix} a & 1-a \\ 1-b & b \end{pmatrix} \qquad (0<a,b<1)。$$

这时定理 5．３９条件满足，故不变分布 $p = \{p_0, p_1\}$ 存在唯一，且应满足下列方程

$$(p_0, p_1)\begin{pmatrix} a & 1-a \\ 1-b & b \end{pmatrix} = (p_0, p_1), \qquad p_0 + p_1 = 1。$$

容易解得

$$p_0 = \frac{1-b}{2-(a+b)}, \quad p_1 = \frac{1-a}{2-(a+b)}。$$

这里我们并不需要知道 $\mathbf{P}^{\mathbf{n}}$ 的形式.

例５．１１（续） 在牙膏品牌的市场占有率的例子里，Markov 链 $\{x_n, n \ge 0\}$ 的所有状态均为遍历态，所以它存在唯一的不变分布 $p = \{p_1, p_2, p_3, p_4\}$，且满足 $p\,\mathrm{P} = p$ ，即

$$0.9p_1 + 0.1p_2 + 0.08p_3 + 0.1p_4 = p_1 ,$$

$$0.05p_1 + 0.8p_2 + 0.1p_3 + 0.1p_4 = p_2 ,$$

$$0.03p_1 + 0.05p_2 + 0.8p_3 + 0.1p_4 = p_3 ,$$

$$0.02p_1 + 0.05p_2 + 0.02p_3 + 0.7p_4 = p_4 ,$$

$$p_1 + p_2 + p_3 + p_4 = 1 .$$

可以求出

$p = \{0.482, 0.253, 0.179, 0.086\}$

可见，随着时间的推移，厂家 A,B,C,D **的市场占有率将分别变为** 48.2%，25.3%，17.9%**和** 8.6% **.**

[**注**] **定理** 5**．３９的条件对于不变分布的存在唯一性来说太强． 事实上，我们有**

（1）**有限状态的** Markov **链只要是互通的，（即全体状态组成一个互通常返类），它一定存在唯一的不变分布．例如，两端反射壁的随机徘徊，它不是一步互通的，**$\mathbf{P}^n$**也没有极限，但是它仍有唯一的不变分布．**

（2）我们将在第 5 节 中得到，可数状态的只有正常返态的 Markov **链，一定存在不变分布．又若此链还是互通的，那么不变分布还是唯一的（但是此时**$\mathbf{P}^n$**未必有极限）．**

3. 3 **转移矩阵的平均极限**

在$n\to\infty$**时，一般地**$\mathbf{P}^n$**未必有极限，例如对于**$\mathbf{P}=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}$，**有** $\mathbf{P}^{2n+1}=\mathbf{P}$，$\mathbf{P}^{2n}=\mathbf{I}$．**极限不存在的原因在于**$\mathbf{P}^n$**有周期性．消除周期性的办法是代之以求平均极限** $\lim_{n\to\infty}\frac{1}{n}(\mathbf{I}+\mathbf{P}+\cdots+\mathbf{P}^{n-1})$．**对于可数状态的** Markov **链，虽然可以证明这个极限一定存在．但是，在本书中我们只对有限状态情形给出证明.**

定理 5．４０ （有限状态情形的平均极限定理） **设**$\mathbf{P}$**为**$N\times N$**转移矩阵，则**

$$\lim_{n\to\infty}\frac{1}{n}(\mathbf{I}+\mathbf{P}+\cdots+\mathbf{P}^{n-1})$$

存在，把它记为 $\mathbf{L}=(l_{ij})$， **则它满足**

$$\mathbf{LP}=\mathbf{PL}=\mathbf{L}=\mathbf{L}^2.$$

证明　记

$$\mathbf{L}(n)\overset{\Delta}{=}\left(l_{ij}(n)\right)=\frac{1}{n}(\mathbf{P}+\mathbf{P}^2\cdots+\mathbf{P}^n).$$

显见 $l_{ij}(n)$ **是取值于**$[0,1]$**的序列，所以存在公共子列** n_k，**使对于任意** $i,j\le N$，$l_{ij}(n)\to$ **某个**l_{ij}．**于是**

$$\mathbf{L}(n_k)\mathbf{P}=\frac{1}{n_k}\sum_{m=1}^{n_k}\mathbf{P}^{m+1}=\frac{1}{n_k}(\sum_{m=1}^{n_k}\mathbf{P}^m+\mathbf{P}^{n_k+1}-\mathbf{P})\to\mathbf{L}.$$

从而有 $\mathbf{LP}=\mathbf{L}$．**同样我们有** $\mathbf{PL}=\mathbf{L}$．**进而得** $\mathbf{LP}^m=\mathbf{L}$，$\mathbf{P}^m\mathbf{L}=\mathbf{L}$ **以及**

$$\mathbf{LL}(n)=\mathbf{L}(n)\mathbf{L}=\mathbf{L}.$$

最后，我们来证明 $\mathbf{L}(n)\to\mathbf{L}$．**为此只要验证** $\mathbf{L}(n)$ **的一切收敛子列都有相同的极限．假**

定另有一个收敛子列：$\mathbf{L}(m_k)\to\tilde{\mathbf{L}}$． 那么，由$\mathbf{L}\mathbf{L}(m_k\}=\mathbf{L}(m_k)\mathbf{L}=\mathbf{L}$ 便得

$\mathbf{L}\tilde{\mathbf{L}}=\tilde{\mathbf{L}}\mathbf{L}=\mathbf{L}$．但是 $\mathbf{L}$ 与 $\tilde{\mathbf{L}}$ 是对称的，因此也应有 $\tilde{\mathbf{L}}\mathbf{L}=\mathbf{L}\tilde{\mathbf{L}}=\tilde{\mathbf{L}}$．这样就有$\tilde{\mathbf{L}}=\mathbf{L}$．

从而 $\mathbf{L}(\mathbf{n})\to\mathbf{L}$．

[注] 对于可数状态的 Markov 链，定理结论仍然正确，但是证明要复杂多了，需要用 Tauber 型定理，故略．

为讨论极限矩阵$\mathbf{L}$的形状，我们先引述一个数学分析的引理，它的证明由数学分析是不难得到的，本书略去．

引理５．４１ 若 $a_n\ge 0,\sum_n a_n<\infty, 0\le b_{n,k}\le M$，且 $b_{n,k}\to b_k,(n\to\infty)$，则

$$\sum_{k=1}^{n}a_k b_{n,k}\xrightarrow{n\to\infty}\sum_{k=1}^{\infty}a_k b_k\ .$$

命题５．４２ (1) $l_{ij}=f_{ij}l_{jj}$．

(2)．对于互通的有限状态 Markov 链，极限矩阵$\mathbf{L}$的各行是一样的，即它具有形式 $\mathbf{L}=\mathbf{1}^{\mathrm{T}}p$，其中$p$ 是$\mathbf{P}$的不变测度．

证明 由

$$l_{ij}(n)=\frac{1}{n}\sum_{m=1}^{n}p_{ij}^{(m)}=\frac{1}{n}\sum_{m=1}^{n}\sum_{k=1}^{m}f_{ij}^{(k)}p_{jj}^{(m-k)}=\sum_{k=1}^{n}f_{ij}^{(k)}\frac{1}{n}\sum_{m=k}^{n}p_{ij}^{(m-k)}\ ,$$

用引理５．４１便得 (1)． 而互通的有限 Markov 链的状态都是常返态，由定理 5．28 可知 $f_{ij}=1$，即 $l_{ij}=l_{jj}$， 从而$\mathbf{L}$具有形式 $\mathbf{1}^{\mathrm{T}}p$，其中$p_j=l_{ij}=l_{jj}$．

注 1 对于可数个状态的互通的常返 Markov 链， 命题也成立，但是p 未必是$\mathbf{P}$的不变测度，因为它可能为零矩阵 (参见后面的 5.2 段)．由定理５.４２可见$\mathbf{L}$应该是以下的形式

$$\begin{array}{cccccc} & T & H_1 & H_2 & \cdots \\ T & \mathbf{0} & f_{ij}p_j^{(1)} & f_{ij}p_j^{(2)} & \cdots \\ \boldsymbol{L}=H_1 & \mathbf{0} & \mathbf{1}^T p^{(1)} & \mathbf{0} & \mathbf{0} \\ H_2 & \mathbf{0} & \mathbf{0} & \mathbf{1}^T p^{(2)} & \mathbf{0} \\ \vdots & \vdots & \vdots & \mathbf{0} & \ddots \end{array}$$

注 2 (有限状态 Markov 链的 f_{ij} 的计算) 由全概率公式可得下述递推关系

$$f_{ij}^{(n)}=\sum_{k\ne j}p_{ik}f_{kj}^{(n-1)},(n>1),\qquad f_{ij}^{(1)}=p_{ij}.$$

对n求和便得

$$f_{ij}=\sum_{k\ne j}p_{ik}f_{kj}+p_{ij}\ .$$

记总状态数为N，于是 $\vec{f_j}\overset{\Delta}{=}(f_{1j},\cdots,f_{Nj})^T$是方程 $\vec{x}=\mathbf{P}^{(j\to 0)}\vec{x}+\overrightarrow{p_j}$ 的解， 其中$\mathbf{P}^{(j\to 0)}$

为把 $\mathbf{P}$ 中第 j 列改为零后所得的矩阵, 而 $\overrightarrow{p_j}$ 为 $\mathbf{P}$ 的第 j 列列向量. 这个方程的近似解可以用迭代方法求得． 用这个方法可以在原则上估算被指定的吸收态所吸收的概率，例如说，从 i 出发最终被吸收态 a 吸收的概率为 f_{ia} ． 因此，上面的方程实际上是博采者输光概率的方程的一般化.

4. Dobrushin 不等式与指数收敛性

研究 $\mathbf{P}^n$ 的收敛性, 比定理 5。３９在理论与应用方面更为一般的是 Dobrushin 方法. 它建立在 Dobrushin 不等式的基础上. 对于无穷维行向量(可取负值) $\mathbf{w} = (w_1, w_2, \cdots)$，可以定义模

$$\|\mathbf{w}\| = \sum_j |w_j|$$

(显见它满足三角不等式: $\|\mathbf{w} + \tilde{\mathbf{w}}\| \le \|\mathbf{w}\| + \|\tilde{\mathbf{w}}\|$).

4. 1　Dobrushin 不等式

定理 5. 4 3　对于概率分布向量 $\mathbf{u}, \mathbf{v}$. 我们有

(1) $\|\mathbf{uP} - \mathbf{vP}\| \ (= \|\sum_i p_{i\bullet} u_i - \sum_i p_{i\bullet} v_i\|) \le C(P)\|\mathbf{u} - \mathbf{v}\|$，

其中

$$C(P) = \frac{1}{2}\sup_{i,k} \sum_j |p_{ij} - p_{kj}| \overset{\Delta}{=} \|\mathbf{p}_i - \mathbf{p}_k\|,$$

称为转移密度的 Dobrushin 收缩系数，简称 Dobrushin 数，其中 $\mathbf{p}_i$ 是 $\mathbf{P}$ 的第 i 行组成的行向量。

(2) 对于转移矩阵 $\mathbf{P}, \mathbf{Q}$ 有

$$C(PQ) \le C(P)C(Q)。$$

证明 (1) 记数 a 的正部与负部分别为 a^+ 与 a^-. 再分别记

$$r_j^{\pm} = (u_j - v_j)^{\pm} / \frac{1}{2}\|\mathbf{u} - \mathbf{v}\|.$$

注意

$$\sum_j [(u_j - v_j)^+ - (u_j - v_j)^-] = \sum_j (u_j - v_j) = 0,$$

$$\sum_j [(u_j - v_j)^+ + (u_j - v_j)^-] = \sum_j |u_j - v_j| = \|\mathbf{u} - \mathbf{v}\|,$$

我们有

$$\sum_j r_j^{+}=\sum_j r_j^{-}=\frac{2}{\|\mathbf{u}-\mathbf{v}\|}\sum_j (u_j-v_j)^{+}=1.$$

同时还有

$$u_j-v_j=(u_j-v_j)^{+}-(u_j-v_j)^{-}=\frac{1}{2}(r_j^{+}-r_j^{-})\|\mathbf{u}-\mathbf{v}\|.$$

于是

$$\|\sum_i p_{i\bullet}u_i-\sum_i p_{i\bullet}v_i)\|=\sum_j \quad |\sum_i p_{ij}(u_i-v_i)|$$

$$=\frac{1}{2}\|\mathbf{u}-\mathbf{v}\|\sum_j \quad |\sum_i p_{ij}(r_i^{+}-r_i^{-})|=\frac{1}{2}\|\mathbf{u}-\mathbf{v}\|\sum_j \quad |\sum_i\sum_k \ (p_{ij}-p_{kj})r_i^{+}r_k^{-}|$$

$$\le\|\mathbf{u}-\mathbf{v}\|\sum_i\sum_k \ r_i^{+}r_k^{-}(\frac{1}{2}|p_{ij}-p_{kj}|)\le C(P)\|\mathbf{u}-\mathbf{v}\|.$$

今证(2). 记P的第i行的行向量为p_i，则由定义及(1)得到

$$C(PQ)=\frac{1}{2}\sup_{i,k}\sum_j|\sum_l \ p_{il}q_{lj}-\sum_l \ p_{kl}q_{lj}|$$

$$=\frac{1}{2}\sup_{i,k}\|\mathbf{p}_i\mathbf{Q}-\mathbf{p}_k\mathbf{Q}\|$$

$$\le\frac{1}{2}\sup_{i,k}C(Q)\|\mathbf{p}_i-\mathbf{p}_k\|=\frac{1}{2}\sup_{i,k}\|=C(P)C(Q).$$?

显见，对于N个状态的Markov链的转移矩阵P有

$$C(P)\le 1-N\min_{i,j}p_{ij}.$$

4.2 Dobrushin 收敛定理

定理5.44 若$C(P)<1$，则存在不变分布p，及常数$0<a<1, C>0, n_0$，使

$$\sum_j|p_{ij}(n)-p_j|<Ca^n,(\forall n\ge n_0).$$

证明提示 对于任意概率分布向量$\mathbf{u},\mathbf{v}$，用 Dobrushin **不等式得**

$$\|\sum_i u_i p_{i\bullet}^{(n)}-\sum_i v_i p_{i\bullet}^{(n)}\|=\|\sum_{i,j}p_{ij}^{(n)}p_{j\bullet}(u_i-v_i)\|$$

$$=\|\sum_j p_{j\bullet}(\sum_i p_{ij}^{(n-1)}u_i-\sum_i p_{ij}^{(n-1)}v_i)\|\le C(p)\|\sum_i p_{ij}^{(n-1)}u_i-\sum_i p_{ij}^{(n-1)}v_i\|$$

$$\le\cdots(\text{归纳地})\cdots\le(C(p))^n\|\mathbf{u}-\mathbf{v}\|.$$

用高等分析中完备距离空间的压缩映射原理（全体向量在距离$d(u,v)\overset{\Delta}{=}\|\mathbf{u}-\mathbf{v}\|$下，组成完备距离空间，参见第9章第3节）可知，存在概率分布向量p，对于任意概率分布向量$\mathbf{u}$，都有

$$\| \sum_i u_i p_{i\bullet}(n) - p \| \le (C(p))^n \| \mathbf{u} - p \|.$$

取$\mathbf{u} = (p_{k1}, p_{k2}, \cdots)$，$C = 2$，$a = C(P)$，便得：对于任意$k$均有

$$\sum_j \mid p_{kj}{}^{(n+1)} - p_j \mid \le Ca^n.$$

[注 1] 定理 5. 3 9 是定理 5.4 4 的特殊情形. 这时有 $C(P) = 1 - Nd < 1$. 但是定理 5.3 9 更为直观, 所以我们仍把它单独列出. 细心的读者会看出, 定理 5.4 4 恰是定理 5.3 9 的发展与抽象化.

[注 2] 在实用中, 定理 5.4 4 大多是适用的.

5. 与常返态相系的延迟更新流, 互通常返 Markov 链的极限定理

5.1 与常返态相系的延迟更新流

对于常返态i，记 Markov 链首次到达i的时刻为T_0，以后各次返回i的时间间隔相继为

$T_1, T_2 \cdots$

命题 5.4 5 $\{T_n : n \ge 1\}$独立同分布，而且与T_0独立.

证明 利用 Markov 性及齐次性

$$P(T_1 = m \mid T_0 = n) = P(x_{n+m} = i, x_{n+l} \ne i, (0 < l < m) \mid x_n = i, x_k \ne i, (0 \le k < n))$$

$$= P(x_{n+m} = i, x_{n+l} \ne i, (0 < l < m) \mid x_n = i) = P(x_m = i, x_l \ne i, (0 < l < m) \mid x_0 = i),$$

而上式右方与T_0的取值无关，因此T_1与T_0独立. 类似地可证明$\{T_n : n \ge 1\}$独立，而且与T_0独立，并且$\{T_n : n \ge 1\}$同分布. ?

从常返态i出发的互通的 Markov 链是一个再现过程. 相系更新间隔流$\{T_n : n \ge 0\}$，由此生成一个延迟更新过程$N_t^{(i)}$. T_1的分布为$d-$格点分布，其中$d \ge 1$. 我们不加证明地指出：在$d \ge 2$时，它就是i的周期.

5. 2 互通常返链的极限定理

定理 5.4 6 设$N_n{}^{(i)}$为互通的 Markov 链在时刻n前返回常返状态i的频数, 则其频率有极限

$$P\left(\frac{N_n{}^{(i)}}{n} \xrightarrow{n \to \infty} \frac{1}{m_i}\right) = 1,$$

其中 $m_i = E(T_1 | x_0 = i) = \sum_k k f_{ii}^{(k)}$，而$\{T_n : n \geq 0\}$为常返状态$i$相系的延迟更新间隔流.

证明　这是延迟更新定理的推论.

[注] 定理的结论也可以写成

$$\frac{I_{\{i\}}(x_1) + \cdots + I_{\{i\}}(x_n)}{n} \xrightarrow{n\to\infty} \frac{1}{m_i} \quad \text{(以概率为 1 成立)}.$$

定理 5.47　若 Markov 链的所有状态都是互通常返的，则对于任意 i 有

$$l_{ij} (= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} p_{ij}^{(k)}) = \frac{1}{m_j},$$

而且此时平均极限$\mathbf{L}$的所有分量要么全为正，要么全为零.

证明　我们只对有限状态情形给出证明，因为一般状态情形需要测度论的工具.
用定理 5.40 我们得到

$$l_{ij} = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} p_{ij}^{(k)} = \lim_{n\to\infty} E(\frac{I_{\{j\}}(x_1) + \cdots + I_{\{j\}}(x_n)}{n} | x_0 = i)$$

$$= E(\lim_{n\to\infty} \frac{I_{\{j\}}(x_1) + \cdots + I_{\{j\}}(x_n)}{n} | x_0 = i) = \frac{1}{m_i}.$$

这里实际上用了取期望与求极限的交换性，在有限状态情形是没有问题的，而在一般情形则需要用测度论的工具.

再则，对于任意两个状态 i, j，由于它们常返互通，就必定存在m_0, n_0，使

$$p_{ij}^{(m_0)}, p_{ji}^{(n_0)} > 0$$

于是

$$l_{ii} = \lim_{n\to\infty} \frac{1}{n + m_0 + n_0} \sum_{k=1}^{n+m_0+n_0} p_{ij}^{(k)}$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} p_{ii}^{(m_0+n_0+k)}$$

$$\geq \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} p_{ij}^{(m_0)} p_{jj}^{(k)} p_{ji}^{(n_0)} = p_{ij}^{(m_0)} l_{jj} p_{ji}^{(n_0)}.$$

又因为 i, j 是对称的，所以它们是否为正是相同的. 从而平均极限$\mathbf{L}$的所有分量要么全为正，要么全为零.

推论 5.48　若 Markov 链的所有状态都是互通正常返的，则$\mathbf{L}$的所有分量全正，此时$\mathbf{L} = \mathbf{1}^{\mathrm{T}} p$，其中 p 是 $\mathbf{P}$ 的唯一的不变测度；而若 Markov 链的所有状态都是互通零常返的，则 $\mathbf{L} = \mathbf{0}$.

推论 5.49　Markov 链存在不变分布的充要条件为至少存在一个正常返类. 在条件成立

下 Markov 链的一切不变分布可以表为它在各个正常返类上的不变分布$p^{(i)}$们的凸线性组合：

$$p = \sum_i l_i p^{(i)},\quad l_i \ge 0, \sum_i l_i = 1.$$

证明提示　由$\mathbf{L}$的形式立得充要条件．用较多的数学工具还可以证明不变分布的表示(因其涉及过多的技术细节，本书略去)．

由推论 2 直接推出下述非常有用的互通正常返性判据．

定理 5．５０ (Markov 链的互通正常返性判据)

（1）对于互通或只有一个常返类的 Markov 链， 那么此链为正常返的充要条件为存在唯一的不变分布．

（2）无暂态的 Markov 链为互通正常返的充要条件为存在唯一的不变分布．　?

这个定理在判别 Markov 链的正常返性时非常有用．

［注 1］当 Markov 链的所有状态都是互通正常返且非周期时，进一步有

$$\lim_{n\to\infty} \mathbf{P}^n = \mathbf{L} = p\ \mathbf{1}^{\mathrm{T}}.$$

此结论在实际中非常有用．它说明在时间发展充分长以后，不论初始值为什么，$\mathbf{P}^n$的每一行都近似为p．它可以用来作为稳态设计的根据.

再则，这个注也是一个非常一般的结论，大多数有关 $\lim_{n\to\infty}\mathbf{P}^n$ 存在的结果，都常由此引申而得.

[注 2] 对于正常返性的判别，还有一个充分条件，我们姑称之为“在有限集外的e－盈函数条件”．近年来，它在多路通信的数学理论中，有不少应用，其叙述如下：

Foster 定理 （利用 Lyapunov 函数的正常返性判据）

若不可约， Markov 链的转移矩阵 P＝(p_{ij}) 满足即： 存在状态空间上的非负函数g (可取＋∞), $C>0$ 及一个状态的有限集F，满足

$$\sum_j p_{ij} g(j) < \infty \qquad (i \in F), \tag{L．1}$$

$$\sum_j p_{ij}(j) g(j) - g(i) \le -C \qquad (i \notin F), \tag{L．2}$$

则此链为正常返的．再加上非周期性，则进一步有

$$\lim_{n\to\infty} \mathbf{P}^n = p\ \mathbf{1}^{\mathrm{T}}.$$

又若（L．2）改为

$$\sum_j p_{ij}(j) g(j) - r\cdot g(i) \le -C \qquad (i \notin F, 0 < r < 1) \tag{L．2）'}$$

那么，此收敛还是指数收敛．

（直观地，函数g与常微分方程的稳定性理论中的 Lyapunov 函数起类似的作用，也称为 Markov 链的 Lyapunov 函数. 第二个条件说明, 函数g在沿F的余集中的状态运动时一致地减小 (经过一次转移后一致地减小了），而由第一个条件，沿有限集F中状态运动时函数g的增加是一致地有界的，且由不可约

性，Markov 链终究会离开有限集 F ，这样就使得在确定极限分布时，F 的作用可以在实际上忽略不计，从而保证极限分布的处在）.

定理的证明需要较多的鞅论知识，本书从略.

更重要的意义是此定理在连续时间的对应的推广形式，称为 Tweedle 定理.

定理 5．51 （遍历性定理）

若 Markov 链的所有状态都是互通正常返的，f 是状态空间 S 上有界实值函数且满足

$\sum_i |f(i)|\frac{1}{\mu_i}<\infty$，则

$$P(\frac{f(x_1)+\cdots+f(x_n)}{n}\xrightarrow{n\to\infty}\sum_i f(i)\frac{1}{m_i})=1，$$

即对于任意一个互通的 Markov 链及满足 $\sum_i |f(i)|\frac{1}{\mu_i}<\infty$ 的 f ，以概率为 1 地有

$$\frac{f(\xi_1)+\cdots+f(\xi_n)}{n}\xrightarrow{n\to\infty}\begin{cases}\sum_i f(i)\pi_i & \text{(对正常返链)}\\ 0 & \text{(对零常返链)}\end{cases}.$$

其中 $p=\{p_1,p_2,\cdots,\}$ 是 P 的不变分布，$p_i=\frac{1}{m_i}$. 此极限与初分布无关.

证明　我们只对有限状态情形给出证明，因为一般状态情形需要测度论的工具．设状态数为 N ．注意到 $f(x_k)=\sum_{i=1}^{N} f(i)I_{\{i\}}(x_k)$，于是由定理 5.40 得

$$\frac{1}{n}\sum_{k=1}^{n} f(x_k)=\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{N} f(i)I_{\{i\}}(x_k)$$

$$=\sum_{i=1}^{N} f(i)\frac{1}{n}\sum_{k=1}^{n} I_{\{i\}}(x_k)\xrightarrow{n\to\infty}\sum_{i=1}^{N} f(i)\frac{1}{m_i}\quad\text{(以概率为 1 成立)}.$$

[注 1]　有多个正常返类时，相应的极限仍然存在，但是依赖于初值的选取.

[注 2] 若 Markov 链 $\{x_n,n\ge 0)$ 的所有状态都是互通正常返，则不难证明 $h_n\overset{\Delta}{=}(x_n,x_{n+1})$ 也是时齐的互通 Markov 链，其转移概率为 $p_{(i,j),(k,l)}=\delta_{kj}p_{jl}$ （它不依赖 i），而且以 $(p_{(i,j)}),(i,j)\in S\times S$ 为唯一的不变分布，其中 $p_{(i,j)}\overset{\Delta}{=}p_i p_{ij}$ ，而 $p=(p_1,p_2,\cdots)$ 是 Markov 链 $\{x_n:n\ge 0)$ 的不变分布．由定理 5.50 推出 $\{h_n:n\ge 0\}$ 也是互通正常返 Markov 链．从而对于 $S\times S$ 上定义的函数 g ，只要满足 $\sum_{i,j}|g(i,j)|\pi_i p_{ij}<\infty$， 就以概率为 1 地

有

$$\frac{g(\xi_1,\xi_{k+1})+\cdots+g(\xi_n,\xi_{k+n})}{n}\xrightarrow{n\to\infty}\sum_{i,j}g(i,j)\pi_i p_{ij}\text{ .}$$

（$h_n \overset{\Delta}{=} (x_n, x_{n+1})$ 的互通性证明如下：

设 $(i,j),(k,l)$ 为任意两个状态。如果 $p_{kl}=0$，则 (k,l) 是一个不可达到的状态，在本质上可以去除。所以我们不妨设 $p_{kl}>0$。由 ξ_n 的互通性，存在一条从 j 到 k 的通路；$j\to j_1\to\cdots\to j_m=k$，于是我们得到：$(i,j)\to(j,j_1)\to(j_1,j_2)\to\cdots\to(j_{m-1},k)\to(k,l)$，这就得到了从 (i,j) 到 (k,l) 的一条通路）.

［注 3］ 一般地，涉及 Markov 链 $\{x_n:n\geq 0)$ 的概率，由它的有限维分布族确定，也就是由它的初始分布 m 与转移矩阵 $\mathbf{P}$ 确定．但是，在得到注 2 的结论时，假定的条件只对 $\mathbf{P}$ 作了要求．这正说明上面的结论与 Markov 链 $\{x_n:n\geq 0)$ 的初始值 m 无关．这类与初始值无关的结论，正适合统计物理中稳态模型的要求（与后面第 11 章的平稳序列理论相比较：当初始值为不变分布 p，即 $P(x_0=i)=p_i$ 时，$\{x_n:n\geq 0)$ 为平稳序列，而且可以证明在我们的条件下，它具有所谓遍历性．于是利用平稳序列的遍历性定理，也可以得到上面的极限．但这时须假定初始值服从不变分布．也就是说，从强有力的平稳序列理论，并不能得到这类与初始值无关的结论．由此可见 Markov 链理论的优点）．

［注 4］注 3 的结论还可以推广到 m 个变量的函数的情形．

［注 5］对于互通的正常返 Markov 链，不管它的初分布 $m=P(x_0=\bullet)$ 是什么，可以用 $\{x_n\}$ 关于时间发展的平均来估计此 Markov 链的不变分布与转移矩阵．作法如下：分别取 $f=I_{\{i\}}, g=I_{\{(i,j)\}}$，则以概率为 1 地有

$$\frac{I_{\{i\}}(x_1)+\cdots+I_{\{i\}}(x_n)}{n}\xrightarrow{n\to\infty}p_i\text{，}$$

$$\frac{I_{\{(i,j)\}}(x_1,x_{k+1})+\cdots+I_{\{(i,j)\}}(x_n,x_{k+n})}{n}\xrightarrow{n\to\infty}p_i p_{ij}(k)\text{ .}$$

由此我们可得用一条轨道来估计 p_i 与 p_{ij} 的估计式：

$$\hat{p}_i=\frac{I_{\{i\}}(x_1)+\cdots+I_{\{i\}}(x_n)}{n}\text{，}$$

$$\hat{p}_{ij}=\frac{I_{\{i\}}(x_1)I_{\{j\}}(x_2)+\cdots+I_{\{i\}}(x_n)I_{\{j\}}(x_{n+1})}{I_{\{i\}}(x_1)+\cdots+I_{\{i\}}(x_n)},$$

此两式是对p_i与p_{ij}进行模拟计算的依据.

关于 Markov 链的数值计算，有以下的注意事项.

1．初始分布$\vec{m}_0$，转移矩阵$\mathbf{P}=(p_{ij})$的 Markov 链的样本，可以作如下的模拟：

(1) 取一个$\vec{m}_0$随机数i_0;

(2) 取一个$\vec{p}_{i_0}\,(=(p_{i_0 1},\cdots,p_{i_0 j},\cdots))$随机数$i_1$;

(3) 取一个$\vec{p}_{i_1}$随机数i_2; …

那么$(i_0,i_1,i_2,\cdots)$就是 Markov 链$(x_0,x_1,x_2,\cdots)$的一个样本.

2．没有可逆分布情形时不变分布的近似算法(参阅[T])

假定状态空间为非负整数，且当$m\ge m_0$时，不变分布$\mathrm{p}=(p_0,p_1,\cdots)$满足

$$p_m\approx Ca^m.$$

那么

$$p_m\approx p_{m_0}a^{m-m_0}\qquad (m\ge m_0).$$

(1) a 的近似确定

此时不变分布的矩母函数$p(z)\approx\sum_{k=0}^{m_0-1}p_k z^k+p_{m_0}z^{m_0}\frac{1}{1-a\cdot z}$. 所以，如果知道了不变分布的矩母函数$p(z)$近似地是一个有理函数，且它的分母只有一个大于 1 的根，那么此根的倒数就是a.

(2) $p_0,p_1,\cdots,p_m$的确定

不变分布满足的方程可以改写为

$$\begin{pmatrix}p_0\\ \vdots\\ p_{m_0-1}\end{pmatrix}=\mathbf{A}\begin{pmatrix}p_0\\ \vdots\\ p_{m_0-1}\\ p_{m_0}\end{pmatrix},$$

$$\sum_{i=0}^{m_0-1}p_i+\frac{p_{m_0}}{1-a}=1,$$

其中

$$\mathbf{A}=(a_{ij})\qquad (0\le i\le m_0-1,0\le j\le m_0),$$

$$a_{ij}=\begin{cases} p_{ji} & (0\le j\le m_0-1) \\ \sum_{k=m_0}^{\infty} a^{k-m_0} p_{ki} & (j=m_0) \end{cases}.$$

6． 停时与强 Markov 性

6. 1 停时

定义５.５２ 非负且可取$+\infty$的随机时刻t，称为时齐的 Markov 链$\{x_n: n\ge 0)$的停时，如果它是$\{x_n\}$-可知的，即：对于任意n，事件$\{t\le n\}$可由 Markov 链$\{x_n\}$在n前的信息$\{x_m: m\le n\}$所决定．直观上说明停时的发展是与 Markov 链的发展是“同步的”.

Markov 链首次到达状态空间的某一个子集合 A 的时刻t_A，就是停时的典例．这是因为事件$\{t_A\le n\}=\bigcup_{m=1}^{n}\{x_m\in A, x_k\notin A, (k<m)\}$只涉及$\{x_n\}$在$n$前的信息．

6.2 强 Markov 性

命题５.５３ 时齐的 Markov 链$\{x_n: n\ge 0)$具有强 Markov 性，就是对于停时也有 Markov 性，即：对于任意停时t及任意状态$i, j, i_0, \cdots, i_{n-1}$，在集合$\{t<\infty\}$上，都有

$$P(x_{t+1}=j\,|\,x_t=i, x_{\bullet}\text{在时刻}t\text{前的信息})=P(x_{t+1}=j\,|\,x_t=i)=P(x_1=j\,|\,x_0=i).$$

(证明提示 把条件概率写成联合概率除以条件的概率，再在计算其中每一个概率时，对t的取值用全概率公式，然后分别用 Markov 性，即可证明．细节从略).

[注] 强 Markov 性也有与 Markov 性类似的相应的等价性叙述，请读者仿照 Markov 性的各种等价性叙述，把它们写出来.

7． 禁忌概率与首达分布

7. 1 禁忌概率

定义５.５４ 由状态i出发，经n步到达状态j，中间未到过状态集A的概率，称为禁忌A的转移概率，记为${}_A p_{ij}^{(n)}$，即

$${}_A p_{ij}^{(n)}=P(x_n=j, x_m\notin A, (0<m<n)\,|\,x_0=i).$$

7. 2 首达时与首达分布

定义５.５５ (首达时及其分布)

设A为一个状态集，随机时刻$t_A\overset{\Delta}{=}\inf(n\ge 0: x_n\in A\}$称为首达$A$的时刻 (对于空集$\Phi$，恒定义

$\inf\ \Phi=\infty$), **简称** A **的首达时.**

(**注意首达时与击中时不同**, **击中时定义为** $T_A \overset{\Delta}{=} \inf(\ n\geq 1: x_n\in\ A\}$, $n=0$ **它在何处是不计入的**).

定义5．56 (**首达分布**)

$P(t_A<\infty, x_{t_A}=j\,|\,x_0=i)$ **作为状态** j **的概率函数，称为** A **的首达分布**. **一般地**, **首达分布是有亏损的分布**, **因为**

$$\sum_{j\in S} P(t_A<\infty, x_{t_A}=j\,|\,x_0=i)=P(t_A<\infty\,|\,x_0=i)\leq 1,$$

而可能小与1．只有当 $P(t_A<\infty\,|\,x_0=i)=1$ **时，首达分布才是一个分布.**

定义5．57 (**越出时**)

$$s_A \overset{\Delta}{=} \inf\{\ n\geq 0: x_n\notin\ A\}$$

称为越出状态集 A **的时刻**, **简称** A **的越出时**. **它是** A **的余集的首达时刻** $t_A=s_{S\setminus A}$.

7．3 **禁忌概率**, **首达分布与平均首达时间**

定理5．58

(1). $${}_A p_{ij}^{(n)}=\begin{cases}P(t_A=n, x_n=j\,|\,x_0=i) & (j\in A)\\ P(t_A>n, x_n=j\,|\,x_0=i) & (j\notin A)\end{cases}.$$

(2). **把** $\mathbf{P}$ **限制在状态集** A **上，并记为** $\mathbf{P_A}=(p_{ij})_{i,\,j\in A}$, **则对于任意** $i, j\notin\ A$ **有**

$${}_A p_{ij}^{(n)}=(\ \mathbf{P_{S\setminus A}}\ ^n)_{ij}.$$

(3) **对于** $\forall i\notin\ A$ **有**

$$P(t_A>n\,|\,x_0=i)=(\mathbf{P_{S\setminus A}}\ ^n\mathbf{1}^T)_i=\sum_{j\notin A} {}_A p_{ij}^{(n)}\ .$$

(4). **若** $P(t_A<\infty)=1$ **，则有**

$$E(t_A\,|\,x_0=i)=(\sum_{\mathbf{n=0}}^{\infty}\mathbf{P_{S\setminus A}}\ ^n\ \mathbf{1}^T)_i\,.$$

以上各式中的上标 n **是指** n **次方幂**.

证明 (1) **显然．今用归纳法证明**(2)**．** $n=1$ **时恰是定义**. **设** n **时等式成立**. $n+1$ **时利用** Markov **性及归纳法假设得**

$$\begin{aligned}{}_A p_{ij}^{(n+1)} &= \sum_{k\in S\setminus A} P(t_A>n, x_n=k, x_{n+1}=j)\\ &= \sum_{k\in S\setminus A} {}_A p_{ik}^{(n)} P(x_{n+1}=j\,|\,x_n=k)\\ &= \sum_{k\in S\setminus A} \mathbf{P_{S\setminus A}}\ ^n p_{kj}=(\ \mathbf{P_{S\setminus A}}\ ^{n+1})_{ij}.\end{aligned}$$

(3)**的证明**:

$$p(t_A > n | x = i) = \sum_{j\in S\setminus A} p_{ij}^{(n)}$$

$$= \sum_{\mathbf{j}\in \mathbf{S}\setminus \mathbf{A}} (\mathbf{P}_{\mathbf{S}\setminus\mathbf{A}}^{\ n})_{ij} = (\mathbf{P}_{\mathbf{S}\setminus\mathbf{A}}^{\ n}\mathbf{1}^T)_i .$$

最后证明(4), **当** $P(t_A < \infty) = 1$**时，我们有**

$$E(t_A | x_0 = i) = \sum_{k=1}^{\infty} kP(t_A = k | x_0 = i) = \sum_{n=0}^{\infty} P(t_A > n | x_0 = i) .$$

再用(3), **便得**(4).

定理 5.5 9 若 $\forall i \in B, P(s_B < \infty | x_0 = i) = 1$, 则有

(1). $$(\mathbf{I} - \mathbf{P_B}^{\mathbf{n}})^{-1} = \sum_{\mathbf{n}=\mathbf{0}}^{\infty} \mathbf{P_B}^{\ n} < \infty .$$

(2) B **的平均越出时间与越出分布有表达公式**:

$$E\ (s_B | x_0 = \bullet) = (\mathbf{I} - \mathbf{P_B})^{-1}\mathbf{1}^T ,$$

$$P(x_{s_\mathbf{B}} = j | x_0 = i) = \sum_{k\in B} [\ (\mathbf{I} - \mathbf{P_B})^{-1}]_{ik}\, p_{kj} .$$

证明 令

$$\tilde{p}_{ij} = \begin{cases} p_{ij} & (i \in B) \\ d_{ij} & (i \in S \setminus B) \end{cases} .$$

那么定理 5．5 8**对于** $\tilde{\mathbf{P}}$ **仍然成立．而对于转移矩阵为** $\tilde{\mathbf{P}}$ **的** Markov **链来说，** $S \setminus B$ **中的状态都是吸收态，由本定理的条件知道** B **中的状态都不是常返态，即全是暂态．故而**

$$\sum_n {}_{S\setminus B}\, p_{ij}^{(n)} = \sum_n \tilde{p}_{ij}^{(n)} < \infty, \qquad \forall i, j \in B .$$

于是由定理 5．5 8**的**(2)**推出本定理的（1）.** (2)**的第一式恰为定理** 5．5 8**的**(4), **而由定理** 5．5 8**的**(1), **有**

$$P(x_{s_\mathbf{B}} = j | x_0 = i) = \sum_{n\geq 1} P(s_B = n, x_n = j | x_0 = i) = \sum_{n\geq 1} {}_{S\setminus B}\, p_{ij}^{(n)} .$$

再用定理 5．5 8**的**(2) **便得第二式．** 】

在 S **为有限集时，平均越出时间与越出分布可写成更为简单的公式，提供实用计算．**

定理 5.6 0 **对于有限状态** Markov **链，平均越出时间与越出分布可以用行列式表示，即对任意** $i \in B$ **只要满足** $P(s_B < \infty | x_0 = i) = 1$, **那么，对任意** $j \in B$ **有**

$$E\ (s_B | x_0 = i) = \sum_{k\in B} \frac{D_{ik}(B)}{D(B)} ,$$

$$(P(x_{s_\mathbf{B}} = j | x_0 = i)) = \sum_{k\in B} \frac{D_{ik}(B)\, p_{kj}}{D(B)} ,$$

其中 $D(B)=\det(\mathbf{I}-\mathbf{P_B})$，$D_{jk}(B)$ 为 $D(B)$ 的 (j,k) 处元素的代数余子式.

证明　把定理５．５９ 的(2)中第一式用逆矩阵写出来就是 (1). 类似地得到第二式.

8. 可逆 Markov 链与可逆分布

一个用 Markov 链描述的系统, 常常可以利用它的长时间后的稳定分布，作为其参数的稳定设计的依据. 在前面我们已经看到, 在很宽的条件下, 不变分布就是这个稳定分布. 而求不变分布需要解一个方程, 一般来说需要很大的工作量. 所幸的是, 在实际中有很多简单的网络系统常满足一个对称性条件, 即所谓可逆性条件．对于这种 Markov 链，有一种求不变分布的简便方法．在本节中，我们将给出这方面的理论．

8.1　可逆 Markov 链

定义５．６１　以 P 为转移矩阵的 Markov 链，称为可逆的，如果存在概率分布 $m=(m_1,m_2,\cdots)$ 使

$$\forall i,j\in S,\qquad m_i p_{ij}=m_j p_{ji}.$$

m 称为可逆初分布，简称可逆分布．在物理中，此条件称为细致平衡条件．

由此定义可以看出，可逆分布是唯一的． 又可逆分布一定是不变分布．事实上，由可逆性的定义立得

$$\sum_i m_i p_{ij}=\sum_i m_j p_{ji}=m_j\ .$$

定义５．６２　随机变量序列 x_n 称为时间可逆的，如果对于任意时刻 $m>n\geq 0$，恒有

$$(x_n,x_{n+1},\cdots,x_m)\text{ 与 }(x_m,\cdots,x_{n+1},x_n)\text{ 同分布．}$$

例５.６３　若取可逆 Markov 链的可逆分布为初始分布，即 $P(x_0=i)=m_i,(\forall i\in S)$，则此 Markov 链是时间可逆的．

事实上，由初始分布是可逆分布得到

$$P(x_n=i_n,x_{n+1}=i_{n+1},\cdots,x_m=i_m)=m_{i_n}p_{i_ni_{n+1}}p_{i_{n+1}i_{n+2}}\cdots p_{i_{m-1}i_m}$$

$$=m_{i_{n+1}}p_{i_{n+1}i_n}p_{i_{n+1}i_{n+2}}\cdots p_{i_{m-1}i_m}=p_{i_{n+1}i_n}p_{i_{n+2}i_{n+1}}\cdots p_{i_mi_{m-1}}m_{i_m}$$

$$=P(x_m=i_m,\cdots,x_{n+1}=i_{n+1},x_n=i_n)\ .$$

命题 5，64　在初始分布 m 下，Markov 链是可逆的充要条件为： 以 m 为初始分布的随机变量序列是时间可逆的．

证明　必要性已经证明. 充分性得自 $\forall i,j\in S$,

$$m_i p_{ij}=P(x_0=i,x_i=j)=P(x_0=j,x_i=i)=m_j p_{ji}.$$

推论５.６５　细致平衡条件等价于: $\forall n,i,j$，有 $m_i p_{ij}^{(n)}=m_j p_{ji}^{(n)}$.

求可逆 Markov 链的不变分布(就是可逆分布)，与求一般 Markov 链的不变分布相比，要简单得多．所以，判别一个 Markov 链是否具有可逆性，在实际中是十分重要的．

再则，若 Markov 链具有不变分布 m．设 S_0 是状态空间的一个子集．把 Markov 链限制在 S_0 上，即对于 $\forall i, j \in S_0$，定义

$$p_{ij}^{S_0} = \frac{p_{ij}}{\sum_{k \in S_0} p_{ik}}, \quad \tilde{p}_{ij} = \begin{cases} p_{ij} & (i \neq j) \\ p_{ii} + \sum_{k \notin S_0} p_{ik} & (i = j) \end{cases},$$

$$m_i^{S_0} = \frac{m_i}{\sum_{k \in} m_k}.$$

一般地，m^{S_0} 不再是 $\mathbf{P}^{S_0}$ 和 $\tilde{\mathbf{P}}$ 的不变分布．但是，当 m 为 $\mathbf{P}$ 的可逆分布时，m^{S_0} 仍是 $\mathbf{P}^{S_0}$ 和 $\tilde{\mathbf{P}}$ 的可逆分布．这个性质称为可逆初始分布关于状态子集的不变性．

8.2 **例**

例５．６６ {整数格点 Z 上的生灭链}

设正数 $p_i, q_i = 1 - p_i, (i = 0, \pm 1, \pm 2, \cdots)$.

$$\mathbf{P} = (p_{ij}), \qquad p_{ij} = \begin{cases} p_i & (j = i+1) \\ q_i & (j = i-1) \\ 0 & (\text{其它}) \end{cases}.$$

定义

$$\tilde{m}_i = \begin{cases} \prod_{k=0}^{i-1} \frac{p_k}{q_{k+1}} & (i > 0) \\ 1 & (i = 0) \\ \prod_{k=i+1}^{0} \frac{q_k}{p_{k-1}} & (i < 0) \end{cases}.$$

那么容易检查

$$\forall i, j \in S, \qquad \tilde{m}_i \, p_{ij} = \tilde{m}_j \, p_{ji}.$$

于是只要 p_i 们满足条件

$$\sum_{i=-\infty}^{\infty} \tilde{m}_i < \infty,$$

再令 $m_i = \dfrac{\tilde{m}_i}{\sum_{j=-\infty}^{\infty} \tilde{m}_j}$，则 $m = (\cdots, m_{-1}, m_0, m_1, \cdots)$ 就是可逆分布．

例 5．67　设转移矩阵为

$$\mathbf{P}=\begin{pmatrix}\frac{1}{2} & \frac{1}{2} & 0\\ 0 & \frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & 0 & \frac{1}{2}\end{pmatrix}.$$

我们来证明它不存在可逆分布. 解方程 $m\mathbf{P}=m$ 得 $\mathbf{P}$ 的不变分布为 $m=(\frac{1}{3},\frac{1}{3},\frac{1}{3})$. 而对于 $i=1,2$ 有 $m_i p_{i,i+1}-m_{i+1}p_{i+1,i}=\frac{1}{6}\neq 0$. 这说明 $\mathbf{P}$ 没有可逆分布.

8. 3　可逆初分布存在性判别法

定理 5.68　(Kolmogorov 可逆性准则)　状态数为 N 的互通的 Markov 链具有可逆分布的充要条件为: 对于任意的一个状态环路 $R: i_1\to i_2\to\cdots\to i_m\to i_1$ 恒成立下面的 (K) 条件：

(K) $$p_R=p_{R^-},$$

其中 p_R 是沿环路 R 的转移概率的积, 即:

$$p_R\overset{\Delta}{=}p_{i_1 i_2}\cdots p_{i_{m-1}i_m}p_{i_m i_1},$$

而 R^- 则是环路 R 的反向环路.

在 (K) 条件满足时, 可逆分布是唯一的, 而且可以求得如下： 固定某个状态 i_0, 对于任意状态 i, 任取一条从 i_0 到的通路 $i_0\to i_1\to\cdots\to i_m\to i_{m+1}=i$. 于是它满足

$$v_i\overset{\Delta}{=}\prod_{k=0}^{m}\frac{p_{i_k i_{k+1}}}{p_{i_{k+1}i_k}}>0.$$

定义

$$m_i=\begin{cases}\dfrac{v_i}{1+\sum_{k\neq i_0}v_k} & (i\neq i_0)\\[2ex] \dfrac{1}{1+\sum_{k\neq i_0}v_k} & (i=i_0)\end{cases}.$$

那么可逆分布为 $m=(m_1,m_2,\cdots,m_N)$，其中 v_i 不依 i_0 及通路的选取.

[注]　对于一般的可数状态 Markov 链, 只要在 (K) 条件外加一个要求 $\sum_i v_i<\infty$, 则充要条件及可逆分布 m 的表达式仍然正确. 可见本定理提供了得到可逆分布的简单方法.

证明　必要性: 首先证明 m_i 们全为正.事实上,由于 m 是可逆初分布, 所以存在 i_0

使 $m_{i_0}>0$. 再由链的互通性, 对于任意 i, 必有 n 使 $p_{i_0 i}^{(n)}>0$. 于是由可逆性条件立得 $m_i>0$. 其次, 任取通路 $i_0\to i_1\to\cdots\to i_{n+1}=i$，则我们有

$$m_{i_0}p_{i_0i_1}p_{i_1i_2}\cdots p_{i_ni}=p_{i_1i_0}m_{i_1}p_{i_1i_2}\cdots p_{i_ni}=\cdots=p_{i_1i_0}p_{i_2i_0}\cdots p_{ii_n}m_i.$$

特别地, 当通路为闭的环路时有 $i_0=i$, 故得条件(K). 因此必要性成立.

充分性：首先证明, v_i 们的定义与从 i_0 到 i 的通路的具体取法无关. 事实上, 设有两条通路: R_1: $i_0\to i_1\to\cdots\to i_{n+1}=i$ 及 R_2: $i_0\to i'_1\to\cdots\to i'_{m+1}=i$. 把沿通路 R_1,R_2 定义的 v_i 值分别记为 $v_i^{(1)}$ 与 $v_i^{(2)}$. 对于 R_1,R_2 的反向环路 R_1^-,R_2^-，考虑复合环路 $R_2^-\circ R_1$ (i_0 出发沿 R_1 至 i 再沿 R_2 的反向环路 R_2^- 回到 i_0), 由 (K) 条件可得 $p_{R_2^-\circ R_1}=p_{R_1^-\circ R_2}$. 即

$$1=\frac{p_{R_2^-\circ R_1}}{p_{R_1^-\circ R_2}}=\frac{p_{R_1}p_{R_2^-}}{p_{R_2}p_{R_1^-}}=\frac{v^{(1)}{}_i}{v^{(2)}{}_i}.$$

于是 $v_i^{(1)}=v_i^{(2)}$，从而 m 作为 **v** 的规一化也是唯一地确定的. 最后证明 m 为可逆分布. 事实上，对于任意的 i,j 分两种情形考虑

(1). $p_{ij}>0$ 的情形. 取 i 到 j 的通路 R. 这时对于前面的通路 R_1，由(K)条件得

$p_{R_1^-\circ R^-}=p_{R\circ R_1}>0$. 可见有 $p_{ji}>0$. 进一步取 $i_0=j$, 那么 $v_i=\dfrac{p_{ji}}{p_{ij}},(i\neq j),v_j=1$.

于是得 $v_ip_{ij}=v_jp_{ji}$. 所以 m 是可逆分布.

(2). $p_{ij}=0$ 的情形. 此时由情形(1)知道必有 $p_{ji}=0$, 从而 $v_ip_{ij}=v_jp_{ji}$ 仍旧成立.

9. 分支 Markov 链 (Galton-Watson 简单分支过程)

Markov 链的应用非常广泛，它涉及物理、生物、化学、神经网络、信息、经济、金融等诸多领域. 在各种应用模式中,由于实际问题往往是十分复杂的，如不对其进行适当的简化,不仅可能使问题因过于复杂而无法解决,而且也会使主要矛盾被枝节所掩盖而难以理解. 因而我们必须对许多实际问题进行必要的抽象和简化，以便理解其实质. 于是，也就不可避免地会使问题可能与实际有一定的差距. 所以，在读者应用 Markov 链去解决实际问题时，还需具体问题具体分析，根据实际情况去作必要的变通，以灵活地应用各种模型.

本节所要介绍的分支 Markov 链，主要关注的是群体灭绝的概率. 这类问题来源于中子，细胞或微生物等(其中个体统一地称为粒子)的分裂与死亡的随机现象. 这里每一代的裂变

是各自独立地进行的，裂变后的粒子的总数决定了下一代粒子的总数的分布．

设$\{x_{n,k}: n,k \ge 1\}$是一族相互独立同分布，而且取非负整数值的随机变量．它表示第 n 代的第 k 个粒子分裂成的粒子数(假定在它分裂后自己就立刻死亡)． $x_{n,k}=0$就意味着此粒子不分裂而死亡．于是第 n＋1 代的粒子总数是$h_{++1}=x_{n,1}+\cdots+x_{n,h_n}$．

设开始是一个粒子，即$h_0=1$．我们说明h_n是 Markov 链．事实上，利用独立性我们有

$$P(h_{n+1}=j\,|\,h_n=i,h_{n-1}=i_{n-1},\cdots)=P(x_{n,1}+\cdots+x_{n,h_n}=j\,|\,h_n=i,h_{n-1}=i_{n-1},\cdots)$$

$$=P(x_{n,1}+\cdots+x_{n,i}=j\,|\,h_n=i,h_{n-1}=i_{n-1},\cdots)=P(x_{n,1}+\cdots+x_{n,i}=j)$$

$$=P(x_{1,1}+\cdots+x_{1,i}=j)\text{，}$$

上式右方与$\{h_{n-1}=i_{n-1},h_{n-2}=i_{n-2},\cdots\}$及$n$无关．可见$h_n$是时齐的 Markov 链，而且其转移概率为

$$p_{ij}=P(x_{1,1}+\cdots+x_{1,i}=j)\text{．}$$

现在我们来具体计算p_{ij}．为此令$p_k=P(x_{1,1}=k)=p_{1k}(k\ge 0)$．记$x_{1,1}$的矩母函数为

$$F(z)=\sum_{k=0}^{\infty}p_k z^k\text{，}$$

那么，利用幂级数的乘法规则，我们由数学归纳法得到$[F(z)]^i=\sum_{k=0}^{\infty}p_{ik}z^k$．于是

$$p_{ik}=\frac{1}{k!}\frac{d[F(z)]^i}{dz}|_{z=0}\quad(i>0,k\ge 0)\text{．}$$

此外，由模型的含义显然有 $p_{0,0}=1, p_{0k}=0$ （$k>0$）．这就是说 0 是此 Markov 链的吸收状态，当过程达到状态 0 就不再有粒子了，分裂也就中止了．在这个模型中，我们首先感兴趣的量是此 Markov 链被 0 吸收的概率，即灭绝概率．我们把它记为 r，即 $r=P(T_0<\infty)$，其中T_0为 Markov 链h_n首次达到 0 的时刻．

记h_n的矩母函数为

$$G_n(z)=\sum_{k=0}^{\infty}P(h_n=k)z^k\text{．}$$

再一次由幂级数的乘法推出

$$G_{n+1}(z)=\sum_{k=0}^{\infty} P(h_{n+1}=k)z^k=\sum_{k=0}^{\infty}\sum_{i=0}^{\infty} P(h_{n+1}=k\,|\,h_n=i)P(h_n=i)z^k$$

$$=\sum_{k=0}^{\infty}\sum_{i=0}^{\infty} p_{ik}P(h_n=i)z^k=\sum_{i=0}^{\infty} P(h_n=i)[F(z)]^i=G_n(F(z)).$$

即第n代的个体数的矩母函数是第 1 代的矩母函数的n次叠置．由此可以算出Eh_n与$Var(h_n)$的递推公式．

现在我们来分别分析在什么条件下有ρ＝1，ρ＝0及0＜ρ＜1．也就是什么条件下粒子概率为 1 地全死光、永远不会死光、或既可能死光也可能不死光．最简单的情况是$p_0=0$或$p_0=1$的情形，它们分别代表粒子在下一个时刻不死亡与必然死亡．那么显然前者永不灭绝，而后者是必定灭绝．为避免这种平凡的情况我们假定0＜ρ＜1． 我们来求ρ．记平均一个粒子的后代个数为$m=Ex_{1,1}$．直观地可以猜测：若μ＞1则不会全部灭绝，而 μ＜1则会全部灭绝． 为此，注意由于各个粒子的分裂是独立地进行的，所以有

$$P(T_0<\infty\,|\,h_1=k)=P(T_0<\infty\,|\,h_1=1)^k=r^k$$

于是由全概率公式推得

$$r=P(T_0<\infty)=p_0+\sum_{k=1}^{\infty} p_kP(T_0<\infty\,|\,h_1=k)=F(r).$$

可见r是函数$F(x)$在$[0,1]$中的不动点．也就是说ρ应为曲线$y=x$与曲线$y=F(x)$在中$[0,1]$之交点的横坐标．显见$x=1$总是一个不动点，这是一个平凡的不动点．我们对此不感兴趣．而当$0<x<1$时有

$$F'(x)=\sum_{n=1}^{\infty} np_nx^{n-1}>0,\ F''(x)=\sum_{n=2}^{\infty} n(n-1)p_nx^{n-2}\ge 0.$$

可见$F(x)$是递增的凸函数．于是$F(x)$是否在区间(0，1)中有不动点，就归结为凸函数$F(x)-x$在$x=1$处的导数是正还是负．前一种情形说明当ε＞0充分小时，$F(1-e)-(1-e)<F(1)-1=0$，因而由$F(0)=p_0>0$与连续函数的中值定理得到: 方程$F(x)=x$在区间$(0,1)$中有解，进而由$F(x)$的凸性就知道这个方程在区间$(0,1)$中解是唯一的，我们把它记为r_0．而后一种情形即$F'(1)-1\le 0$（也即μ≤1），此时由$F(x)$的凸性得到$F(x)-x$在区间$(0,1)$中恒负．所以这时方程$F(x)=x$在[0，1]中只可能有一个解:

$x=1$. 从而在 $m<1$时必然有 $r=1$, 这说明最终将以概率为 1 地全部灭绝. 而在μ＞1时,

我们来证明 $r=r_0$ (而不可能是 $r=1$). 事实上

$$r=\sum_{n=1}^{\infty} f_{1,0}^{(n)}=\lim_{n\to\infty}\sum_{k=1}^{n} f_{1,0}^{(k)} p_{00}^{(n-k)}=\lim_{n\to\infty} p_{1,0}^{(n)}.$$

为了证明 $r=r_0$, 我们只需归纳地证明 $p_{10}^{(n)}\le r_0$. 首先, 当$n=1$时有

$$p_{1,0}^{(1)}=p_0=F(0)<F(r_0)=r_0.$$

今作归纳法假设: 设此不等式在n时成立. 在$n+1$时由$F(x)$的递增性我们有

$$p_{1,0}^{(n+1)}=\sum_{k=1}^{\infty} p_{1k}^{(1)} p_{k0}^{(n)}=\sum_{k=1}^{\infty} p_k (p_{1,0}^{(n)})^k=F(p_{1,0}^{(n)})\le F(r_0)=r_0.$$

综上所述得到以下结论

定理5．69　对于 Galton-Watson 简单分枝过程, 设一个粒子的裂变分布列为 $P(x_{1,1}=k)=p_k$ （k＝0，1，…）, 且满足$0<p_0<1$. 令 $m=Ex_{1,1}$, 则在开始时有 i 个粒子时, 全部粒子的灭绝概率为 r^i, 其中

$$r=\begin{cases}1 & (m\le 1)\\ r_0 & (m>1)\end{cases},$$

而 r_0是方程$F(x)=x$的最小非负解.

此时第n代粒子数的平均规模为

$$a_{n+1}\overset{\Delta}{=}E(h_{n+1}\mid h_0=1)=\sum_{k=1}^{\infty} kp_{1k}^{(n+1)}=\sum_{k=1}^{\infty} k\sum_{j=1}^{\infty} p_j p_{jk}^{(n)}$$

$$=\sum_{j=1}^{\infty} p_j E(h_n\mid h_0=j)=\sum_{j=1}^{\infty} p_j\, jE(h_n\mid h_0=1)=ma_n.$$

由于$a_0=1$, 因而$a_n=m^n$. 可见当μ＜1时, 平均粒子数逐代单调下降趋于 0, 当μ＝1 时, 各代粒子平均相同, 而当μ＞1时, 平均粒子数按指数阶逐代上升至无穷. $m=1$的情形称为临界情形, 前面已经知道在这种情形粒子最终必定灭绝, 此时人们感兴趣的问题是平均灭绝时间多长, 对此本书不再介绍. 而$m>1$的情形, 则称为上临界情形.

习题 5

1. 举例说明，如果集合 A 包含不止一个状态，那么对于 Markov 链 $\{X_n: n\geq 0\}$，下列等式不一定成立

$$P(X_{n+1}=j \mid X_0=i_0,\cdots,X_{n-1}=i_{n-1},X_n\in A)=P(X_{n+1}=j \mid X_n\in A).$$

2. 设 x_n 是以 $\{0,1,2\}$ 为状态空间的 Markov 链，其转移矩阵为 $\begin{pmatrix} \frac{a}{2} & 1-a & \frac{a}{2} \\ \frac{a}{2} & \frac{a}{2} & 1-a \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$. 问

$h_n = I_{\{2\}}(x_n)$ 是否为一个 Markov 过程？如果回答"是"，则转移矩阵是什么？

3. 设有数字 1，2，…，m.

(1). 在其中任取一个数记为 X_0. 再在 1，2，…，X_{n-1} $(n\geq 1)$ 中任取一个数记为 X_n. 证明 $\{X_n: n\geq 0\}$ 为 Markov 链，并求其转移矩阵.

(2). 把它作一个随机排列：$a_1,\cdots,a_m$. 令 $X_0=1$，$X_n=\min\{i>X_{n-1}, a_i\geq a_k (k<i)\}$，$\min\varnothing$ 定义为 $m+1$. 证明 $X_0,\cdots,X_{m-1}$ 是一个以 $\{1,\cdots,m+1\}$ 为状态空间的 Markov 链. 求它的转移矩阵

4. 连续地掷一枚均匀的硬币. 如果第 n 次掷得正面，则令 $X_n=0$；如果第 n 次掷得反面，且从第 n 次往前看连续出现反面的数目为 k，则令 $X_n=k$. 约定 $X_0=0$. 证明 $\{X_n: n\geq 0\}$ 是 Markov 链，并求其转移矩阵.

5. 汽车单向地通过某路口. 假设一秒钟可通过一辆汽车. 路口的红灯持续 r 秒，绿灯持续 g 秒. 以"红灯-绿灯"的出现作为一个单元时间. 在第 n 个单元开始时在等候的汽车数记为 X_{n-1}，而把在这单元中来到的汽车数记为 x_n. 设 $\{x_n: n\geq 1\}$ 为独立同分布 $x_n \sim \begin{pmatrix} 0 & 1 & \cdots & r+g \\ p_0 & p_1 & \cdots & p_{r+g} \end{pmatrix}$. 再假定在开始时刻路口没有汽车在等候. 证明 $\{X_n: n\geq 0\}$ 是 Markov 链，并求其转移矩阵.

6. 设 m 个人中的某些人已患流行性感冒. 假定传染在两人间的接触中发生，当病人遇见健康者时，后者被传染的概率为 p，一切成对的接触都是等可能的，而且在每个单位时间内只发生一次接触. 记 X_n 为时刻 n 患病的人数. 证明 X_n 是 Markov 链，并求其转移矩阵.

7. 设某水库容量为 c 立方米. 在时刻 n 水库的储水量记为 X_n，在时段 $[n,n+1]$ 内流入水库的水量记为 x_{n+1}.

超过水库容量的水即溢出. 假定在时段$[n,n+1]$末从水库放掉 m（$<c$）立方米水（其中包括库满而溢出的部分，如果溢出的水量已超过 m 立方米就不再放水）. 假设 $\{x_n : n\ge 1\}$ 独立同分布且与 X_0 独立，$P(x_n = k) = a_k, k\ge 0$. 证明 $\{X_n : n\ge 0\}$ 是 Markov 链, 并求其转移矩阵.

8. 设 $p,q>0, p+q=1$. 若 Markov 链具有循环型的转移矩阵:

$$P=\begin{pmatrix} q & p & & & & \\ & q & p & & & \\ & & \ddots & \ddots & & \\ & & & & & \\ & & & & q & p \\ p & & & & & q \end{pmatrix}.$$

证明

$$P^n=\begin{pmatrix} a_1^{(n)} & a_2^{(n)} & \cdots & a_{m-1}^{(n)} & a_m^{(n)} \\ a_m^{(n)} & a_1^{(n)} & \ddots & & a_{m-1}^{(n)} \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & a_2^{(n)} \\ a_2^{(n)} & \cdots & & a_m^{(n)} & a_1^{(n)} \end{pmatrix},$$

其中 $a_1^{(n)}+a_2^{(n)}w+\cdots+a_m^{(n)}w^{m-1}=(q+pw)^n$，而 w 为 1 的 m 次主复数根(1 的其它复根都是它的幂).

9. 证明有限状态 Markov 链的转移矩阵的特征值的模都不大于 1, 而且 1 总是特征值.

10. 对于二值 Markov 链

$$P=\begin{pmatrix} a & 1-a \\ 1-b & b \end{pmatrix}, (0\le a,b\le 1)$$

求 $f_{ij}^{(n)}$ 及 $\lim_{n\to\infty} P^n$.

11. 对下列各转移矩阵的状态进行分类:

$$P_1=\begin{pmatrix} & & \frac{1}{2} & \frac{1}{2} \\ 1 & & & \\ & 1 & & \\ & 1 & & \end{pmatrix},\ P_2=\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & & & \\ \frac{1}{2} & \frac{1}{2} & & & \\ & & \frac{1}{2} & \frac{1}{2} & \\ & & \frac{1}{2} & \frac{1}{2} & \\ \frac{1}{4} & \frac{1}{4} & & & \frac{1}{2} \end{pmatrix},\ P_3=\begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix},$$

$$P_4=\begin{bmatrix}\frac13&\frac23&0\\\frac12&0&\frac23\\0&\frac13&\frac23\end{bmatrix},\ P_5=\begin{bmatrix}0&\frac13&\frac23\\\frac13&0&\frac23\\\frac13&\frac23&0\end{bmatrix},\ P_6=\begin{pmatrix}1&&&&\cdots\\\frac12&\frac12&&&\cdots\\\frac13&&\frac23&&\cdots\\\frac14&&&\frac34&\cdots\\\vdots&&&&\ddots\end{pmatrix}.$$

对 P_4,P_5,P_6 求 f_{ij}. 再记从状态 1 出发首次返回状态 1 的时间为 t_1, 求它的分布与期望.

12．设 Markov 链的转移矩阵为 $P=\begin{bmatrix}\frac12&\frac14&\frac14\\0&\frac34&\frac14\\0&0&1\end{bmatrix}$， 求 $f_{ii},(i=1,2,3)$ 及 $\lim_{n\to\infty}P^n$.

13. 若出口商品用三状态表示： 状态为+1 表示今年比去年增长, 状态 0 表示与去年持平, 状态 −1 表示 今年比去年减少．过去的统计数据得到的状态间的转移矩阵为

$$\begin{array}{c}\\-1\\0\\+1\end{array}\begin{array}{c}\begin{matrix}-1&0&+1\end{matrix}\\\begin{bmatrix}0&0.2&0.8\\0.35&0.3&0.35\\0.6&0.4&0\end{bmatrix}\end{array}$$

求各状态的平均返回时间和增长趋势与减少趋势的期望长度.

14．证明定理５．３９是定理５．４４的特例．

15． 对于时齐的 Markov 链 $\{x_n:n\ge0\}$．令 $h_n=(x_n,x_{n+1})$，则 h_n 也是时齐的 Markov 链．在求 h_n 的转移矩阵．

16 证明对于互通的有限 Markov 链 $\{x_n:n\ge0\}$，以概率为 1 地有极限关系

$$\frac{g(x_1,x_{k+1})+\cdots+g(x_n,x_{k+n})}{n}\xrightarrow{n\to\infty}\sum_{i,j}g(i,j)p_i p_{ij}(k)$$

17. 设 $\{X_n:n\ge0\}$ 是为正常返非周期的不可约 Markov 链，转移矩阵为 $P=(p_{ij})$，而 $\{x_n:n\ge0\}$ 是与 $\{X_n:n\ge0\}$ 独立的独立同分布.随机变量列，$P(x_n=k)=a_k,(k\ge1)$. 令

$$h_0=0,h_n=x_1+\cdots+x_n\ n\ge1，Y_n=X_{h_n},n\ge0,$$

则 $\{Y_n:n\ge0\}$ 为正常返不可约马尔可夫链， 其转移矩阵为 $\tilde{P}=\sum_{k=1}^{\infty}a_kP^k$. 且它与 $\{X_n:n\ge0\}$ 有相同的平稳分布.

18.. 水库供水按其水位分为下列 5 个状态: 1 表示危险水平, 2 表示缺水, 3 表示刚够, 4 表示达到较好水平, 5 表示充足. 由统计已有数据得到其转移矩阵为

$$P=\begin{bmatrix}0.1 & 0.1 & 0.3 & 0.5 & 0\\ 0.3 & 0.2 & 0.2 & 0.2 & 0.1\\ 0.1 & 0.2 & 0.4 & 0.2 & 0.1\\ 0 & 0.1 & 0.2 & 0.4 & 0.3\\ 0 & 0.1 & 0.1 & 0.4 & 0.4\end{bmatrix}.$$

求到出现危险水平的平均时间.

19. 设t 为 Markov 链x_n的有限值停时，令$h_n=x_{t+n}$. 证明h_n也是 Markov 链，求它的转移矩阵. 再证明在已知$h_0=x_t$的条件下，$\{h_n\}$与$(t,x_0,\cdots,x_t)$条件独立.

20. 设$\{x_n\}$，$\{h_n\}$相互独立，它们分别是独立同分布序列$\{X_n)$与$(Y_n\}$的部分和：$x_n=X_1+\cdots+X_n$，$X_k\sim\begin{pmatrix}0 & 1\\ q_1 & p_1\end{pmatrix},(k\ge 1)$，$h_n=Y_1+\cdots+Y_n$，$Y_k\sim\begin{pmatrix}0 & 1\\ q_2 & p_2\end{pmatrix},(k\ge 1)$. 令 $T=\min\{n:n\ge 1,x_n-h_n=\pm M\}$. 证明

（1）若$p_1>p_2$，则$P(x_T-h_T=-M)=\dfrac{1}{1+l^M}$，其中 $l=\dfrac{p_1q_2}{p_2q_1}$；

（2）$ET=\dfrac{M(l^M-1)}{(p_1-p_2)(1+l^M)}$.

21. 设 $\{x_n:n\ge 0\}$ 独立同分布，$x_n\sim\begin{pmatrix}-1 & 1\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}$，$S_0=0,S_n=\sum_{k=1}^{n}x_k,n\ge 1$，$M_n=\max\limits_{0\le k\le n}S_k,n\ge 0$. 证明$\{|S_n|:n\ge 0\}$及$\{M_n-S_n:n\ge 0\}$都是 Markov 链.

22. 假定$\{x_n,n\ge 1\}$独立同分布.，$P(x_n=k)=a_k,(k\ge 1)$. 对于$n\ge 0$定义

$$h_n=\begin{cases}0, & n<x_1\\ k, & x_1+\cdots+x_k\le n<x_1+\cdots+x_{k+1},k\ge 1\end{cases}，V_n=n-\sum_{j=0}^{h_n}x_j,(x_0=0).$$

证明$\{V_n:n\ge 0\}$是常返的不可约 Markov 链. 而且它是正常返的充要条件为$\sum_{k=1}^{\infty}ka_k<\infty$.

23. 设 Markov 链x_n有N个状态，而i为常返态. 求证存在常数q，$0<q<1$，使得当$n>N$时，从状态i出发首次返回i的时间t_i满足$P(t_i>n\mid x_0=i)<q^n$.

24. 证明对于状态空间为S的 Markov 链有

(1) 对于任意$i,j\in S$，有$f_{ij}^{(1)}=p_{ij}$，$f_{ij}^{(n+1)}=\sum\limits_{k\ne j}p_{ik}f_{kj}^{(n)},(n\ge 1)$.

(2) 在 j 固定时，$\{f_{ij}:\forall i\in S\}$ 是无穷个线性方程组 $x_i=\sum_{k\neq j}p_{ik}x_k+p_{ij}$ 以初值为 $\{x_i^{(0)}=0,\forall i\in S)$ 的迭代解 $\{x_i^*=0,\forall i\in S)$，即 $x_i^*=\sum_n x_i^{(n)},(\forall i\in S)$，而 $x_i^{(n+1)}=\sum_{k\neq j}p_{ik}x_k^{(n)}+p_{ij}$.

(3) 上面的迭代解 $\{x_i^*=0,\forall i\in S)$ 是方程 $x_i=\sum_{k\neq j}p_{ik}x_k+p_{ij}$ 的最小非负解.

25. 设取值于非负整数的 Markov 链的状态为 $\{0,1,2,\cdots\}$， 转移矩阵为

$$P=\begin{pmatrix} 1 & & & & & \\ (1-d)q & d\cdot q & p & & & \\ & & q & p & & \\ & & & q & p & \\ & & & & \ddots & \ddots \\ & & & & & \end{pmatrix}.$$

求吸收概率 $f_{i0},i\geq 1$.

26. 袋中有黑白两色的球共 m 个. 每次从中随机地取出一个球后，放进一个不同色的球. 令 x_n 为取了 n 次后袋中的白球数. 证明它是正常返不可约 Markov 链，并求其平稳分布.

27. 有 m 个白球与 m 个黑球混放在两个袋中，每个袋中各 m 个. 在每个袋中各取出一个，互相交换后放回袋中. 令 x_n 为交换了 n 次后在第一个袋中的白球数. 证明它是正常返不可约 Markov 链，并求其平稳分布.

28. 若 Markov 链 x_n 具有双随机的转移矩阵，即转移矩阵还满足 $\sum_i p_{ij}=1$. 证明

(1) 如果状态只有有限个，那么它们都是正常返的.

(2) 如果状态只有有限个，且链是不可约的，那么它具有均匀的平稳分布.

(3) 如果状态空间为无限，且链是不可约的，那么它不是正常返的.

29. 设 Markov 链正常返不可约，其平稳分布为 $\mathrm{p}=(p_0,p_1,\cdots)$. 定义倒逆转移矩阵为 $\hat{p}_{ij}=\frac{p_j}{p_i}p_{ji}$. 证明它也是不可约正常返的.

30. 若 Markov 链 x_n 的转移矩阵满足：存在非负数列 u_n 及一个状态 k 使

$$d\overset{\Delta}{=}\sum_j p_{kj}u_j<\infty,\ \sum_j p_{ij}u_j\leq u_i-1,(i\neq k).$$

(1) 作序列 $u_i^{(1)}=u_i,u_i^{(n+1)}=\sum_i p_{ij}^{(n)}u_j$，证明

$$u_i^{(n+2)}\leq(1+d)p_{ik}^{(n)}-1+u_i^{(n+1)}.$$

(2) 证明 $\frac{1}{n}\sum_{m=1}^{n} p_{ik}^{(m)} \geq \frac{1}{1+d}(1-\frac{u_i^{(2)}}{n})$.

(3) 如果x_n还是不可约的，那么它是正常返的.

31. 设 Markov 链的转移矩阵为 $\begin{pmatrix} a_0 & a_1 & a_2 & \cdots \\ 1 & & & \\ & 1 & & \\ & & \ddots & \end{pmatrix}$ ($a_i > 0, i \geq 0$). 证明它是常返不可约.，而且它是正常返的充要条件为 $m = \sum_{n=1}^{\infty} na_n < \infty$. 在条件成立时平稳分布为 $\text{p} = (p_0, p_1, \cdots)$，其中 $p_i = \frac{1}{m}\sum_{n=i}^{\infty} a_n (i = 0,1,\cdots)$.

32. 设 Markov 链的转移矩阵为$\begin{pmatrix} q_0 & 1-q_0 & & \\ q_1 & & 1-q_1 & \\ q_2 & & & \ddots \\ \vdots & & & \end{pmatrix}$($0 < q_i < \infty, \quad i = 0,1,\cdots$). 证明

(1) 它是非常返的充要条件是$\sum_i q_i < \infty$;

(2) 它是正常返的充要条件是$\sum_i q_i = \infty$，且$m = 1 + \sum_{I=1}^{\infty} (1-q_0)\cdots(1-q_{i-1}) < \infty$. 在条件成立时平稳分布为 $\text{p} = (p_0, p_1, p_2, \cdots) = \frac{1}{m}(1, 1-q_0, (1-q_0)(1-q_1), \cdots)$.

33. 设 Markov 链的转移矩阵满足

$$p_{i,i-1} = q_i > 0, \quad p_{ii} = r_i \geq 0, \quad p_{i,i+1} = p_i > 0 \quad (i = \cdots, -1, 0, 1, \cdots)$$

给出具有平稳分布的条件，并求出其平稳分布.

34. 设x_n是不可约的且具有平稳分布的 Markov 链，而$\{X_n\}$是与此 Markov 链独立的独立同分布随机序列，且 $x_n \sim \begin{pmatrix} 1 & \cdots & k & \cdots \\ p_1 & \cdots & p_k & \cdots \end{pmatrix}$. 记$h_n = x_{X_1+\cdots+X_n}$， 证明 h_n 也是不可约正常返 Markov 链，并求它的转移矩阵，再说明它与x_n有相同的平稳分布.

35. 设$h_{n+1} = x_{n,1} + \cdots + x_{n,h_n}$为分支链，满足$P(x_{1,1} = k) = p_k \, (k \geq 0)$.

(1) 假定$p_0 + p_1 + p_2 = 1$，证明此分支链以概率为 1 灭绝的充要条件为 $p_0 \geq p_2$.

(2) **假定** $p_0+p_1+p_2+p_3=1$，**证明此分支链以概率为** 1 **灭绝的充要条件为** $p_0 \ge 2p_3+p_2$.

(3) **假定** $p_3=1$，**求此分支链以灭绝的概率**.

(4) **假定** $p_k=ar^k(k\ge 1, 0<r<1, 0<a\le 1-r)$，**求此分支链以灭绝的概率**.

(5) **假定** $p_0=1-p, p_1=p$，**求此分支链灭绝时间的分布**.

龚光鲁，钱敏平著 应用随机过程教程及其在算法与智能计算中的应用

清华大学出版社,2003

第6章 连续时间的 Markov 链 (Q-过程)

1 连续时间的 Markov 链及其转移矩阵

1．1 连续时间的 Markov 链的定义及等价性描述

定义６．１ 随机变量族$\{x_t : t \ge 0\}$称为连续时间的 Markov 链，如果这些随机变量都是离散的 (状态空间S至多是一个可数集，即它或者有限，或者与自然数一一对应)，而且对于$\forall m \ge 0, \forall s > s_1 > \cdots s_m \ge 0$及任意状态$i, j, i_1, \cdots, i_m$，都有

$$P(x_{t+s} = j \mid x_s = i, x_{s_1} = i_1, \cdots, x_{s_m} = i_m) = P(x_{s+t} = j \mid x_s = i). \qquad (6.1)$$

Markov 性可以有许多等价叙述，我们概括如下：

等价叙述１ 若A为只与资料$\{x_u, u < s\}$有关的一个事件，则有

$$P(x_{s+t} = j | x_s = i, A) = P(x_{s+t} = j | x_s = i). \qquad (6.\ 2)$$

等价叙述 2 对于过程在时刻s以后所确定的事件B及等价叙述 1 中之A有

$$P(B | x_s = i, A) = P(B | x_s = i). \qquad (6.\ 3)$$

等价叙述 3 在已知“现在”的条件下,“将来”与“过去” 是条件独立的.

等价叙述 4 对状态空间上的任意有界函数f，及$s > s_1 > \cdots > s_m > 0$，均有

$$E(f(x_{s+t}) | x_s = i, x_{s_1} = i_1 \cdots, x_{s_m} = i_m) = E(f(x_{s+t}) | x_s = i) \qquad (6.\ 4)$$

((6. 1)式也是 (6. 4)式的特例，即 $f(x) = I_{\{j\}}(x)$ 的情形).

等价性质 5（最一般的形式） 对于常见的实数集合Λ，只由随机序列$\{x_t\}$在时刻s及其后的信息所决定的随机变量h，以及任意$s > s_1 > \cdots > s_m > 0$， 恒有

$$P(h \in \Lambda | x_s = i, x_{s_1} = i_1 \cdots, x_{s_m} = i_m) = P(h \in \Lambda | x_s = i)$$

或更一般地

$$E(h | x_s = i, x_{s_1} = i_1 \cdots, x_{s_m} = i_m) = E(h | x_s = i). \qquad (6。5)$$

1. 2 连续时间的 Markov 链概率转移矩阵

定义６．２ 记

$$p_{ij}(s,t) = P(x_t = j \mid x_s = i), (t \ge s).$$

定义无穷矩阵

$$\mathbf{P}(s,t)=(p_{ij}(s,t)) ,$$

称为转移矩阵．补充定义 $\mathbf{P}(s,s)=\mathbf{I}$ (无穷单位矩阵).

命题６.３ （概率转移矩阵族的性质）

与离散时间情形类似地，对于为分量全是 1 的无穷行向量(矩阵) 1,我们有

(P.1) $0\le p_{ij}(s,t)\le 1$，$\mathbf{P}(s,t)\mathbf{1}^{\mathrm{T}}=\mathbf{1}^{\mathrm{T}}$.

(P.2) (Chapman-Kolmogorov 方程) 对于$\forall u\ge t\ge s$有

$$\mathbf{P}(s,u)=\ \mathbf{P}(s,t)\ \ \mathbf{P}(t,u), \tag{6. 6}$$

其分量形式为

$$p_{ij}(s,u)=\sum_k p_{ik}(s,t)p_{kj}(t,u). \tag{6. 6)'}$$

证明与离散时间情形类似.

１．３ 连续时间的时齐的 Markov 链

定义６．４ 连续时间的 Markov 链称为时齐的, 如果其转移阵 $\mathbf{P}(s,s+t)$ 与 s 无关．此时我们另记

$$\mathbf{P}(t-s)=\mathbf{P}(s,t)。 \tag{6. 7).}$$

那么 Chapman-Kolmogorov 方程变为

$$\mathbf{P}(s+t)=\ \mathbf{P}(s)\ \ \mathbf{P}(t) \tag{6. 8}$$

在本书中, 除非特别声明，我们所考虑的连续时间的 Markov 链均为时齐的．与时间离散的情形类似, 连续时间的 Markov 链的转移矩阵 $\mathbf{P}(t)$ 是刻画此链的统计特征的要素．即有

定理 6．５ （Markov 链的有限维分布与绝对概率）

(1) 若时续的时间 Markov 链$\{x_t,t\ge 0\}$的初始分布为 $m_i(0)=P(x_0=i)$，其转移矩阵为 $\mathbf{P}(t)=(p_{ij}(t))$，则对于$\forall 0<t_1<\cdots<t_n$，有

$$P(x_0=i_0,x_{t_1}=i_1,\cdots,x_{t_n}=i_n)=m_{i_0}(0)p_{i_0i_1}(t_1)p_{i_1i_2}(t_2-t_1)\cdots p_{i_{n-1}i_n}(t_n-t_{n-1}) \tag{6. 9}$$

(2) 设 $m_i(t)=P(x_t=i)$，并记由 $m_i(t)$ 构成的行向量为 $m(t)=(m_i(t):i\in S)$，那么

$$m(t+s)=m(s)\ \mathbf{P}(t)\text{，从而有}\quad m(t)=m(0)\ \mathbf{P}(t)$$

从而连续时间的 Markov 链的统计性质（包括其长期行为）完全由转移矩阵 $\mathbf{P}(t)$ 及初始分布 $m(0)$ 所决定. 】

连续时间的 Markov 链也有强 Markov 性.

离散时间的时齐 Markov 链的多步转移矩阵 $\mathbf{P}^{(n)}$ 是由一个 ”最小的” $\mathbf{P}$ 所生成的, 而在

连续时间时，在 $\mathbf{P}(t)$ 中找不到直接生成它的"最小者". 但是，从 Chapman-Kolmogorov 方程

$\mathbf{P}(s+t)=\mathbf{P}(s)\mathbf{P}(t)$，$\mathbf{P}(0)=\mathbf{I}$ 的形式看，它是一个无穷维的矩阵值函数方程. 我们拿一个与它相象而最且简单的函数方程 $f(t+s)=f(t)f(s), f(0)=1$ 分析，这个方程在 $t=0$ 处连续的通解为且而解 $f(t)$ 则完全由 $a=f'(0)$ 确定. 由此猜测，在正常的情形下，如果 $\mathbf{P}(t)\xrightarrow{t\to 0}\mathbf{I}$ 满足，$\mathbf{P}(t)$ 在 $t=0$ 处的微商 $\mathbf{P}'(0)$ 似乎应该存在，而且描述 $\mathbf{P}(t)$ 的演化进程的最基本的量也应该就是 $\mathbf{P}'(0)$.

事实上，在实际应用中都可以无妨地认为，$\mathbf{P}'(0)$ 确实有限，而且唯一确定了 $\mathbf{P}(t)$ (用严格的数学可以证明，$\mathbf{P}'(0)$ 只是在较广的意义理解下存在，即：$p_{ij}'(0),(i\neq j)$ 存在，但 $p_{ii}'(0)$ 却可能取 $-\infty$，而且这样的 $\mathbf{P}'(0)$ 未必能唯一地确定 $\mathbf{P}(t)$，这些都是理论概率论中的纯数学问题，而在实际应用中一般是不会遇到的). 于是就应用范围所关心的视点而言，我们可以说，在实用中遇到的连续时间的 Markov 链的转移矩阵 $\mathbf{P}(t)$ 关于 t 是可微的，而且 $\mathbf{P}(t)$ 可由它在 $t=0$ 处的微商 $\mathbf{P}'(0)$ 所确定.

定义6.6 设 $\mathbf{P}(t)\xrightarrow{t\to 0}\mathbf{I}$ 满足，把 $\mathbf{P}'(0)$ 记为 $\mathbf{Q}$，称为转移矩阵 $\mathbf{P}(t)$ 的转移速率矩阵或形式生成元（有时也简称为 $\mathbf{Q}$ 矩阵） 即

$$\mathbf{Q}\overset{\Delta}{=}\mathbf{P}'(0). \tag{6.10}$$

概率速率矩阵 $\mathbf{Q}$ 之所以重要，是因为一般 $\mathbf{P}(t)$ 不能直接由测量得到，然而 $\mathbf{Q}$ 却可通过实验手段测得. 由它再通过解一个称为 Kolmogorov 方程的矩阵微分方程就可解出 $\mathbf{P}(t)$ (参见后文). 这样就能得到连续时间的 Markov 链的统计分布. 这构成了研究连续时间的 Markov 链的统计行为的主要思路.

又由于形式生成元 $\mathbf{Q}$ 与转移矩阵 $\mathbf{P}(t)$ 间的密切关系，所以在我国的概率界常称 $\mathbf{P}(t)$ 为转移速率矩阵 $\mathbf{Q}$ 的 Q-过程. 这是一个具有中国特色的称谓.

在本章中，我们恒假定满足

$$\mathbf{P}(t)\xrightarrow{t\to 0}\mathbf{I}. \tag{6.11}$$

2 Poisson 过程与复合 Poisson 过程再访

例6.7 （Poisson 过程作为 Markov 过程转移矩阵与转移速率矩阵）

Poisson 过程是独立增量过程，所以它有 Markov 性．故它是连续时间的 Markov 链，由第３章知道 Poisson 过程的转移矩阵依赖于连续时间参数t， 即

$$\mathbf{P}(t)=(p_{ij}),\qquad p_{ij}(t)=\begin{cases} e^{-lt}\dfrac{(lt)^{j-i}}{(j-i)!} & (j\ge i) \\ 0 & (\text{其它情形}) \end{cases}.$$

因此转移矩阵是一个上三角形无穷矩阵．这时转移速率矩阵（形式生成元）为

$$\mathbf{Q}=(q_{ij})=\begin{pmatrix} -l & l & 0 & & & \\ 0 & -l & l & 0 & & \\ & 0 & -l & l & 0 & \\ & & & \ddots & \ddots & \ddots \\ & & & & \ddots & \ddots & \ddots \\ & & & & & & \end{pmatrix}.$$

注意 Poisson 过程的转移速率矩阵$\mathbf{Q}$的每一行的和都是 0. 这个事实对于只有有限个状态的连续时间的Markov链总是对的，写成向量形式就是$\mathbf{Q}\,\mathbf{1}^{T}=0$．但是对于一般有可数个状态的连续时间的 Markov 链确未必正确． 因为在数学上我们只能证明 $\mathbf{Q}\,\mathbf{1}^{T}\le 0$， 因此也就引出了许多的困难，从而激起了纯理论研究的概率论学者的兴趣． 好在实际应用中的连续时间的 Markov 链， 常常都满足$\mathbf{Q}\,\mathbf{1}^{T}=0$．

定义６．８ 满足条件

$$\mathbf{P}(t)\xrightarrow{t\to 0}\mathbf{I},\quad \mathbf{Q}\,\mathbf{1}^{T}=0 \qquad (6．12)$$

的$\mathbf{Q}$称为保守的转移速率矩阵．

在实际应用中就忽视 $\mathbf{Q}\,\mathbf{1}^{T}\ne 0$ 的情形．这就是说，在实际应用中遇到的转移矩阵总认为是保守的。 即如果发现了非保守情形，则可能是遗漏了某些状态．

例６．９（复合 Poisson 过程的转移矩阵与转移速率矩阵）

设N_t为强度 l 的 Poisson 过程， $\{Y_k\}$为与之独立的独立同分布随机序列， 则

$$x_t \overset{\Delta}{=} Y_1+\cdots+Y_{N_t}$$

是复合 Poisson 过程．又因为它也是独立增量过程，所以它有 Markov 性．假定

$$Y_k\sim\begin{pmatrix} 1 & 2 & \cdots & j & \cdots \\ f_1 & f_2 & \cdots & f_j & \cdots \end{pmatrix},\ (\sum_j f_j=1).$$

那么x_t是连续时间的 Markov 链．我们来求它的转移矩阵与转移速率矩阵．对于 $s<t$

$$p_{ij}(s,t)=0,(i>j),\quad p_{ii}(s,t)=P(N_t-N\quad=0)=e^{-l(t-s)},$$

而当 $i < j$ **时， 由独立增量性及全概率公式得到**

$$p_{ij}(s,t) = P(x_t = j \mid x_s = i) = P(Y_{N_s+1} + \cdots + Y_{N_t} = j - i \mid Y_1 + \cdots + Y_{N_t} = i)$$

$$= P(Y_{N_s+1} + \cdots + Y_{N_t} = j - i) = P(Y_1 + \cdots + Y_{N_t - N_s} = j - i)$$

$$= \sum_{k \ge 1} P(Y_1 + \cdots + Y_k = j - i \mid N_t - N_s = k) P(N_t - N_s = k)$$

$$= \sum_{k \ge 1} P(Y_1 + \cdots + Y_k = j - i) P(N_t - N_s = k) 。$$

归纳地记概率向量 $\boldsymbol{f} = (f_1, \cdots, f_i, \cdots)$ **的** k **次卷积(它们代表** k **个独立同分布的随机变量的和的分布)为**

$$f^1(l) = f_l, f^{*k}(l) = \sum_{j=1}^{l} f^{*k-1}(j) f^1(l-j) .$$

那么我们得到

$$p_{ij}(s,t) = \sum_{k \ge 1} f^{*k}(j-i) \cdot e^{-l(t-s)} \frac{(l(t-s))^k}{k!} .$$

这说明转移矩阵是时齐的上三角形无穷矩阵．易见其转移速率矩阵为

$$\mathbf{Q} = (q_{ij}) = \begin{pmatrix} -l & lf_1 & lf_2 & \cdots & \cdots & & \\ 0 & -l & lf_1 & lf_2 & \cdots & \cdots & \\ & 0 & -l & lf_1 & lf_2 & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & \ddots & \ddots & \ddots \\ & & & & & & \end{pmatrix} . \qquad (6.13)$$

由此可见复合 Poisson 过程的 $\mathbf{Q}$ **矩阵也是保守的．**

3． 由转移速率矩阵确定时间连续的 Markov 链

3. 1 Kolmogorov 方程及 Master 方程

定理６．１０

(1) $\mathbf{Q} = \mathbf{P}'(0) = (q_{ij})$ **在广义下存在，即**:

$$-\infty \le q_{ii} \le 0, q_{ij} \ge 0, (i \ne j), \sum_j q_{ij} \le 0 . \qquad (6.14)$$

(2) **若总状态数有限， 则** $q_{ii} > -\infty$，**且** $\mathbf{Q}$ **保守 (即** $\sum_j q_{ij} = 0$). **此时** $\mathbf{P}(t)$ **由** $\mathbf{Q}$ **唯一地确定，而且它可以表达为以下的收敛级数**:

$$\mathbf{P}(t) = e^{\mathbf{Q}t} \overset{D}{=} \mathbf{I} + \frac{\mathbf{Q}}{1!} + \frac{\mathbf{Q}^2}{2!} + \cdots; \qquad (6.15)$$

另一方面，$\mathbf{P}(t)$ 也可以通过求解下面两个方程之一得到：

Kolmogorov 向前方程（Fokker-Plank 方程）

$$\mathbf{P}'(t)=\mathbf{P}(t)\,\mathbf{Q}，\qquad \mathbf{P}(0)=\mathbf{I}\,; \tag{6.16}$$

Kolmogorov 向后方程: $\mathbf{P}'(t)=\mathbf{Q}\,\mathbf{P}(t)$, $\quad \mathbf{P}(0)=\mathbf{I}$ (6.17)

（3）对于非有限个状态的 Markov 链，只要$\mathbf{Q}$保守（即$\mathbf{Q}\,\mathbf{1}^T=0$），且 $q_i \overset{\Delta}{=} -q_{ii}$ 是 i 的有界函数，则 (2) 中的所有结论全都成立.

（本定理的证明冗长，且需要较多的数学准备知识，本书中从略．我们只指出：Kolmogorov 向前方程与 Kolmogorov 向后方程可以由 Chapman-Kolmogorov 方程 $\mathbf{P}(t+s)=\mathbf{P}(t)\,\mathbf{P}(s)$ 及 $\mathbf{P}(t+s)=\mathbf{P}(s)\,\mathbf{P}(t)$ 形式地对 s 取微商后再令 $s=0$ 得到).

［注］ 应用中的绝大多数的转移速率矩阵，都满足本定理中(3)的条件，最典型的情形出现在从每个状态只能转移到有限个状态的情形．

从本段以下，我们恒假定如下条件满足：

$$\lim_{t\to 0} p_{ij}(t)=d_{ij}，\ q_i \text{ 有界，}\mathbf{Q}\text{保守（}\mathbf{Q}\,\mathbf{1}^T=0\text{）}. \tag{6. 18}$$

具有满足(6.18)的转移速率矩阵$\mathbf{Q}$的转移矩阵$\mathbf{P}(t)$，称作以$\mathbf{Q}$为转移速率矩阵的连续时间的 Markov 链．.

Kolmogorov 向后方程写成分量就是

$$p'_{ij}(t)=-q_i p_{ij}(t)+\sum_{k\neq i} q_{ik}\,p_{kj}(t)\,.$$

把它看成 $p_{ij}(t)$ 的微分方程，乘上其积分因子为 $e^{q_i t}$ 后积分便得

$$e^{q_i t} p_{ij}(t) = p_{ij}(0)+\int_0^t \sum_{k\neq i} e^{q_i s} q_{ik}\,p_{kj}(s)ds\,.$$

于是得到下面的结论．

定理 6．1 1 （Kolmogorov 向后方程的分量形式的积分迭代）

Kolmogorov 向后方程就等价于积分方程组

$$p_{ij}(t)=e^{-q_i t}d_{ij}+\int_0^t \sum_{k\neq i} e^{-q_i s} q_{ik}\,p_{kj}(t-s)ds\,.$$

由这个方程可作如下的迭代算法：

$$p_{ij}^{(0)}=0$$

$$\cdots$$

$$p_{ij}^{(n)}(t)=e^{-q_i t}d_{ij}+\int_0^t \sum_{k\neq i} e^{-q_i s} q_{ik}\,p_{kj}^{(n-1)}(t-s)ds\,.$$

当 $n\to\infty$ 时，$p_{ij}^{(n)}$ 收敛，其极限就是 Kolmogorov 向后方程的解．

（此定理的证明不难，读者可自行验证）

3．2 转移速率矩阵的概率含义

我们引述下面的定理，它在已知概率速率矩阵Q时，对于模拟对应的连续时间的 Markov 链的轨道，起关键的作用.

定理6．１2

设转移速率矩阵Q满足（6．18）的连续时间的 Markov 链为x_t．记t 为次 Markov 链首次跳跃的时刻， 即

$$t = \inf\{ t : x_t \neq x_0 \}\,.$$

那么

(1) $P(t \le t \mid x_0 = i) = 1 - e^{-q_i t}$ （即：在条件概率$P(* \mid x_0 = i)$下，$t \sim \exp_{q_i}$）.

(2) t 与 x_r 在条件概率$P(* \mid x_0 = i)$下条件独立.

(3) 在条件概率$P(* \mid x_0 = i)$下， ． $x_t \sim \begin{pmatrix} \ldots & i-1 & i+1 & \cdots & \cdots \\ \ldots & \dfrac{q_{i,i-1}}{q_i} & \dfrac{q_{i,i+1}}{q_i} & \ldots & \ldots \end{pmatrix}$.

（在定理假定下，Markov 链的轨道是右连续的， 而定理结论的证明，需要对轨道作细致的分析，本书从略）.

定理6．１３　设条件（6。18）成立．记 t_k 为 Markov 链x_t 的第 k 次跳跃时刻，则 $h_k \overset{\Delta}{=} x_{t_k}$ 是离散时间的时齐 Markov 链， 其转移矩阵为 $\tilde{\mathbf{P}} = (\tilde{p}_{ij})$ ，其中 $\tilde{p}_{ij} = \begin{cases} \dfrac{q_{ij}}{q_i} & (j \neq i) \\ 0 & (j = i) \end{cases}$ ，h_k 称为连续时间的Markov链x_t的嵌入链,它表达了Markov链x_t 的跳跃情况.

（证明提示： 只需检查 Markov 性）.

4．连续时间的 Markov 链的极限分布

4．1　连续时间的 Markov 链的转移矩阵的平均极限

与定理5.４０ 完全类似地可以得到

定理6．１4　满足条件（6．18）的连续时间的 Markov 链的转移矩阵有平均极限

$$\frac{1}{T}\int_0^T \mathbf{P}(t)dt \xrightarrow{T\to\infty} \text{某个}\mathbf{L}\,.$$

而且它满足

$$\mathbf{P}(t)\,\mathbf{L} = \mathbf{L}\ \mathbf{P}(t) = \mathbf{L} = \mathbf{L}^2\,.$$

同时L 具有离散时间情形类似的分块形式.

4．2　连续时间的 Markov 链的极限分布

引理6.１5 设条件（6．18）满足，则

(1) 对于任意 i 及任意$t \ge 0$，恒有 $p_{ii}(t) > 0$.

(2) 对于 $j \neq i$，要么 $p_{ij}(t) \equiv 0$；要么存在t_0，使$p_{ij}(t) > 0, (\forall t \geq t_0)$.

证明　我们只对有限状态情形证明.

(1) 由于 $\mathbf{P}(t) \xrightarrow{t \to 0} \mathbf{I}$，必定存在$h_0 > 0$，当$h \leq h_0$时，对于任意$i$，有 $p_{ii}(h) > 0$. 从而对于任意t，有 $p_{ii}(t) \geq p_{ii}(\frac{t}{n})^n > 0$.

(2)对于$i \neq j$，若$p_{ij}(t) \equiv 0$不成立，则必定存在t_0，使$p_{ij}(t_0) > 0$. 于是由(1)得到

$$p_{ij}(t) \geq p_{ij}(t_0) p_{jj}(t - t_0) > 0, (\forall t \geq t_0).$$

定理6．１６　对于满足(6.18)条件的连续时间的Markov链，存在行和为１的矩阵$\mathbf{P}(\infty)$，使

$$\mathbf{P}(t) \xrightarrow{t \to \infty} \mathbf{P}(\infty).$$

证明　我们也只对有限个状态的情形证明. 由引理6.15(1)，对任意$h > 0$，Markov链 $\{x_{nh} : n \geq 0\}$ 所有的状态i都是非周期的. 因为i是常返态，由定理5．50注1推得$\lim_{n \to \infty} p_{ii}(nh)$存在. 我们把它记成 $p_{ii}^{(h)}(\infty)$.

对于$i \neq j$，若$p_{ij}(t) \equiv 0$不成立，则由引理6.1(2) $\exists t_0$，使$p_{ij}(t) > 0, (\forall t \geq t_0 > 0)$.对于$h \geq t_0$ ，因为i是常返态且i可达j，从而i与j属于同一个非周期常返类. 由定理5.50注1推得$\lim_{n \to \infty} p_{ij}(nh)$存在. 总之有 $\mathbf{P}(nh) \xrightarrow{n \to \infty} \mathbf{P}^{(h)}(\infty)$.

再则，由定理6. 14，$\mathbf{P}(nh)$的平均极限$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \mathbf{P}(kh)$也应为$\mathbf{L}$，从而$\mathbf{P}^{(h)}(\infty)$=$\mathbf{L}$.

最后，我们来证明$\mathbf{P}(t) \xrightarrow{t \to \infty} \mathbf{L}$. 事实上，由于$\mathbf{P}(t) \xrightarrow{t \to 0} \mathbf{I}$，当$h$小时有$|p_{kj}(h) - d_{kj}| < e$，又对于任意充分大的 t，有 $t - [\frac{t}{h}]h < h$，因而

$$|p_{ij}(t) - p_{ij}([\frac{t}{h}]h)| \leq \sum_k p_{ik}([\frac{t}{h}]h) \, |p_{kj}(t - [\frac{t}{h}]h) - d_{kj}| < e.$$

由此可见

$$\lim_{t \to \infty} p_{ij}(t) = \lim_{t \to \infty} p_{ij}([\frac{t}{h}]h) = l_{ij}.$$

引理6．１７　对于满足（6。18）的连续时间的Markov链$\{x_t : t \geq 0\}$，如果i, j关于它

的嵌入链 $\tilde{P}=(\tilde{p}_{ij})$ 互通（也称为 $\mathbf{Q}$ －互通），其中 $\tilde{p}_{ij}=(1-d_{ij})\frac{q_{ij}}{q_i}$， 则对于任意 h，状态 i, j 关于 Markov 链 $\{x_{nh}:n\geq 0\}$ 也是互通的.

证明　我们只对有限状态情形给出证明.

由互通可知存在 $i_1,\cdots,i_m$，使 $i_k\neq i_{k+1},(k=0,\cdots,n;i_0=i,i_{n+1}=j)$，$q_{ii_1}q_{i_1i_2}\cdots q_{i_nj}>0$.

注意到当 $h\leq$ 某个 h_0 时，由 $\mathbf{P}(t)\xrightarrow{t\to 0}\mathbf{I}$，对于任意 $i\neq j$ 有 $p_{ij}(h)=q_{ij}h+o(h)$，便得到 $p_{ii_1}(h)p_{i_1i_2}(h)\cdots p_{i_nj}(h)>0$. 即对于任意 $h\leq h_0$，i 与 j 关于 Markov 链 $\{x_{nh}:n\geq 0\}$ 是互通的.

对于 $h>h_0$ 的情形，只要注意对于 h 给定，必然存在一个 m 使 $\frac{h}{m}<h_0$，由此对于任意的 i, j，只要 $p_{ij}(\frac{h}{m})>0$，利用引理５．１５就得 $p_{ij}(h)\geq p_{ij}(\frac{h}{m})p_{jj}(\frac{h}{m})^{m-1}>0$.　　】

定理 6．１８　对于满足（6．18）条件的连续时间的 Markov 链 $\{x_t:t\geq 0\}$，如果其转移速率阵为 $\mathbf{Q}$ －互通（ 意即对于任意 $i\neq j$，存在 $i_1,\cdots,i_m$，使

$$i_k\neq i_{k+1},(k=0,\cdots,n;i_0=i,i_{n+1}=j),\quad q_{ii_1}q_{i_1i_2}\cdots q_{i_nj}>0\text{）},$$

那么

$$\mathbf{P}(t)\xrightarrow{t\to\infty}=\mathbf{1}^T p\ ,$$

此时 p 或者是一个概率分布，或者恒为０.

证明　由引理 6.１７，对 $\forall h$，$\{x_{nh}:n\geq 0\}$ 都是互通的 Markov 链. 由定理 6．１６可知此时 $\mathbf{L}$ 具有相同的行向量故有形式 $\mathbf{L}=\mathbf{1}^T p$.

5．连续时间的 Markov 链的转移矩阵 $\mathbf{P}(t)$ 的不变分布

5．1　连续时间的 Markov 链的转移矩阵 $\mathbf{P}(t)$ 的不变分布与其嵌入链的不变分布

定义６．１９　概率分布向量 p 称为 $\mathbf{P}(t)$ 的不变分布，如果对于任意 t 恒有

$$p=p\ \mathbf{P}(t)\,.$$

在(6.18)条件满足下，它还等价于

$$p\ \mathbf{Q}=0\,.$$

事实上，我们只在有限状态情形给出证明. 等价的必要性显然. 今用后向方程证明充分性. 利用

$$(p\mathbf{P}(t))' = p(\mathbf{P}(t))' = p\mathbf{Q}\mathbf{P}(t) = 0,$$

便得到 $p\mathbf{P}(t) = p\mathbf{P}(0) = p$.

命题６．２０ 以 $\mathbf{P}(t)$ 为转移矩阵，并以其不变分布 p 为初始分布的连续时间的 Markov 链 $\{x_t : t \ge 0\}$，是平稳过程，即：对于任意 $s,k,t_1,\cdots,t_k$，$(x_{t_1+s},\cdots,x_{t_k+s}\}$ 与 $(x_{t_1},\cdots,x_{t_k}\}$ 有相同的联合分布.

定义６．２１ 概率分布向量 $\tilde{p}$ 称为嵌入链的转移矩阵 $\tilde{\mathbf{P}}$ 的不变分布，如果 $\tilde{p} = \tilde{p}\tilde{\mathbf{P}}$.

命题 6．２２ 若 p 为连续时间的 Markov 链的转移矩阵 $\mathbf{P}(t)$ 的不变分布．则 $\tilde{p} = (\tilde{p}_1,\cdots,\tilde{p}_i,\cdots)$ 是嵌入链 $\tilde{\mathbf{P}}$ 的不变分布，其中 $\tilde{p}_i = \dfrac{p_i q_i}{\sum_j p_j q_j}$；反之，若 $\tilde{p}$ 是嵌入链的转移矩阵 $\tilde{\mathbf{P}}$ 的不变分布，且 $Z \overset{\Delta}{=} \sum_j \dfrac{\tilde{p}_j}{q_j} < \infty$，则 $p = (p_1,\cdots,p_i,\cdots)$ 是 $\mathbf{P}(t)$ 的不变分布，其中 $p_i = \dfrac{\tilde{p}_i}{q_i Z}$.

(请读者自己验证)

我们不加证明地给出下面的定理.

5．2 连续时间的 Markov 链的遍历极限

定理６．２３ 在 (6．18)条件下，若存在唯一的概率分布 p 满足 $p\mathbf{Q} = 0$，则

$$\mathbf{P}(t) \xrightarrow{t\to\infty} = \mathbf{1}^T p \ .$$

此时还有

（1） 对于定义在状态空间上的函数 $f(i)$，只要满足 $\sum_i |f(i)| p_i < \infty$，就以概率为 1 地有

$$\frac{1}{t}\int_0^t f(x_s)ds \xrightarrow{t\to\infty} \sum_i f(i) p_i \ .$$

（2）对于定义在状态空间上的函数 $g(i,j)$，只要满足 $\sum_i |g(i,j)| p_i p_{ij}(u) < \infty$，就以概率为 1 地有

$$\frac{1}{t}\int_0^t g(x_s, x_{s+u})ds \xrightarrow{t\to\infty} \sum_i g(i,j) p_i p_{ij}(u) \ .$$

(多元函数也有类似的结论．得到这个定理的过程与离散时间的 Markov 链的情形类似)

［注］（关于禁忌概率）

与离散时间的 Markov 链类似地，对于有限状态的连续时间的 Markov 链，记

$$ {}_A p_{ij}(t) \overset{\Delta}{=} P(t_A > t, x_t = j \mid x_0 = i), (i, j \notin A), $$

再记 Q 在状态集 A 上的限制为 $\mathrm{Q}_{\mathbf{A}}$，即: $\mathrm{Q}_{\mathbf{A}} = (q_{ij})_{i,j\in A}$, $\mathrm{P}(t)$ 在状态集 A 上的限制为 $\mathrm{P}_A(t)$，

即: $\mathrm{P}_{\mathbf{A}}(t) = (p_{ij}(t))_{i,j\in A}$. 那么可以证明, 对于 $\forall i, j \notin A$ 有

$$ \mathrm{P}_{\mathbf{A}}(t) = e^{t\mathbf{Q}_A}. $$

5．3． 对称的与可逆的连续时间的 Markov 链

定义６．２４ 转移速率矩阵 $\mathbf{Q}$ 称为配称的，如果存在非负数列向量 $m = (m_1, \cdots, m_i, \cdots)$ 满足

$$ m_i q_{ij} = m_j q_{ji}. \tag{6．19} $$

m 称为 $\mathbf{Q}$ 的配称列．易见，m 称为 $\mathbf{Q}$ 的配称列等价于 $(m_1 q_1, \cdots, m_i q_i, \cdots)$ 为嵌入链的转移矩阵 $\tilde{\mathbf{P}}$ 的配称列，即：对于任意 $i \neq j$，恒有 $m_i q_i \tilde{p}_{ij} = m_j q_j \tilde{p}_{ji}$． 连续时间的 Markov 链的转移矩阵 $\mathbf{P}(t)$ 称为对称的，如果存在非负数列向量 $m = (m_1, \cdots, m_i, \cdots)$，对任意 t 满足

$$ m_i p_{ij}(t) = m_j p_{ji}(t), \tag{6．20} $$

（也称此连续时间的 Markov 链是对称的）而 m 称为 $\mathbf{P}(t)$ 的配称列； $\mathbf{P}(t)$ 称为可逆的，如果它有一个配称列 p 为概率分布向量，这时 p 称为可逆分布（也称此连续时间的 Markov 链是可逆的）．可逆分布必是不变分布．

命题６．２５ 在（6．18）条件下，连续时间的 Markov 链的转移矩阵 $\mathbf{P}(t)$ 为对称的，且以 p 为其可逆分布，当且仅当，它的转移速率矩阵 $\mathbf{Q}$ 以 p 为配称列．

在处理一些实际问题中， 下面的一个定理相当好用：

定理 6.２６ 若（6．18）满足，且 $\mathbf{Q}$ 以概率分布向量 p 为其配称列，且其转移速率矩阵 $\mathbf{Q}$ －互通，则

$$ \mathbf{P}(t) \xrightarrow{t\to\infty} = \mathbf{1}^T p . $$

定理 6.２７ (Kolmogorov 对称性准则)

$\mathbf{Q}$ 互通的的连续时间的 Markov 链为对称的充要条件为： 对于任意的一个状态环路 $R: i_1 \to i_2 \to \cdots \to i_m \to i_1$ 成立下面的 (K) 条件：

(K) $$ q_R = q_{R^-}, $$

其中 q_K 是沿环路 R 的转移概率速率的积, 即

$$q_R \overset{\Delta}{=} q_{i_1 i_2} \cdots q_{i_{m-1} i_m} q_{i_m i_1},$$

而 R^- 则是环路 R 的反向环路.

在 (K) 条件满足时, 配称列 $m=(m_1,\cdots,m_i,\cdots)$ 可以求得如下： 固定某个状态 i_0, 对于任意状态 i, 任取一条从 i_0 到的通路 $i_0 \to i_1 \to \cdots \to i_m \to i_{m+1} = i$.令

$$v_i \overset{\Delta}{=} \prod_{k=0}^{m} \frac{q_{i_k i_{k+1}}}{q_{i_{k+1} i_k}} > 0 .$$

定义

$$m_i = \begin{cases} \dfrac{v_i}{1+\sum\limits_{k \neq i_0} v_k} & (i \neq i_0) \\ \dfrac{1}{1+\sum\limits_{k \neq i_0} v_k} & (i = i_0) \end{cases} .$$

在 $\{q_i\}$ 非有界时，条件（6。18）不成立，此时下面引述的定理对 Q 过程的存在性给出了一种可能性

定理 6.２８　若转移速率阵Q是保守的（即行和为 1)，又假定它有配称列 m，满足：

$$\sum_i m_i < \infty，\ \sum_i q_i m_i < \infty .$$

那么此时 $\mathbf{P}(t)$ 由 Q 唯一地确定, $\mathbf{P}(t)$ 可以通过求解 Kolmogorov 向前方程:

$\mathbf{P}'(t) = \mathbf{P}(t)\mathbf{Q}, \mathbf{P}(0) = \mathbf{I}$; 或 Kolmogorov 向后方程: $\mathbf{P}'(t) = \mathbf{Q}\mathbf{P}(t), \mathbf{P}(0) = \mathbf{I}$

得到, 而且 $\mathbf{P}(t)$ 以 $p = \dfrac{1}{\sum_i m_i} m$ 为可逆分布.

(此定理见之于作者之一的工作 <平稳马氏链的可逆性>, 北京大学学报, 第 4 期, 1-9 页, 1978).

6. 例

6.1　连续时间分支过程

假定一个粒子的各次分裂时间组成一个强度为 l 的指数流，各个粒子的分裂是相互独立的. 每个粒子每次分裂时自己消失并产生子代，各个粒子各次分裂后产生的子代的粒子数都是独立同分布的随机变量，其分布为

$$\begin{pmatrix} 0 & 1 & \cdots & k & \cdots \\ f_0 & f_1 & \cdots & f_k & \cdots \end{pmatrix}.$$

令 X_t 为时刻 t 的粒子数，我们来说明它是 Markov 链. 记

$$p_{ij}(t,t+h)=P(X_{t+h}=j\mid X_t=i).$$

我们分析如下：

1． 在$(t,t+h]$中，一个粒子分裂两次以上的概率记为a，则$a=o(h)$．

事实上，在$(t,t+h]$中第一次分裂的时刻必在某个时刻$t+s$，而在$(t+s,t+h]$中至少还要分裂一次．由于这两个是相互独立的，故它们都发生的概率应为$le^{-l(t+s)}\cdot(1-e^{-l(h-s)})$．于是由(推广的)全概率公式得

$$a=\int_0^h le^{-l(t+s)}\cdot(1-e^{-l(h-s)})ds=e^{-lt}(e^{-lh}-1)-e^{-lt}lhe^{-lh}=o(h).$$

2． 在$(t,t+h]$中，一个粒子不分裂的概率为e^{-lh}；而分裂一次的概率为lhe^{-lh}（这是因为在$(0,t]$中的分裂次数是 Poisson 过程)．

3． i个粒子在$(t,t+h]$中至少有两个分裂的概率记为b，则$b=o(h)$．

事实上，

$$b=\sum_{2\le k\le i}C_i^k(lhe^{-lh})^k(e^{-lh})^{i-k}=o(h).$$

概括起来，在$(t,t+h]$中，如忽略$o(h)$不计，那么或者没有粒子分裂，或者只能有一个粒子分裂，而且它也只能分裂一次． 于是我们有

(1)

$$p_{ii}(t,t+h)=P((t,t+h]\text{内}i\text{个粒子全不分裂}\mid X_t=i)$$

$$+P((t,t+h]\text{内}i\text{个粒子中恰有一个分裂一次且分裂为一个}\mid X_t=i)$$

$$=(e^{-lh})^i+C_i^1lhe^{-lh}f_1(e^{-lh})^{i-1}=(1-ilh)+ilhf_1+o(h).$$

从而以下极限与t无关：

$$q_{ii}(t)\overset{\Delta}{=}\lim_{h\to0}\frac{p_{ii}(t,t+h)-1}{h}=-il(1-f_1).$$

(2)对于$j\ge i+1$，

$p_{ij}(t,t+h)$

$$=P((t,t+h]\text{内}i\text{个粒子中恰有一个分裂一次且分裂为}j-(i-1)\text{个}\mid X_t=i)$$

$$=C_i^1lhe^{-lh}f_{j-(i-1)}(e^{-lh})^{i-1}=ilhf_{j-i+1}+o(h).$$

从而以下极限也与t无关：

$$q_{ij}(t)\overset{\Delta}{=}\lim_{h\to 0}\frac{p_{ij}(t,t+h)}{h}=i\,l f_{j-i+1}\,.$$

（3）

$p_{i,i-1}(t,t+h)=P((t,t+h]$内$i$个粒子中恰有一个分裂一次且分裂为0个$\mid X_t=i)$.

故而以下极限也与t无关：

$$q_{i,i-1}(t)\overset{\Delta}{=}\lim_{h\to 0}\frac{p_{i,i-1}(t,t+h)}{h}=i\,l f_0\,.$$

（4）当$j<i-1$时，$p_{ij}(t,t+h)=0(h)$，所以$q_{i,j}(t)\overset{\Delta}{=}\lim_{h\to 0}\dfrac{p_{i,j}(t,t+h)}{h}=0$

即X_t具有与t无关的概率转移矩阵：

$$\mathbf{Q}=\begin{pmatrix} 0 & 0 & & & & \\ l\cdot f_0 & -l\cdot(1-f_1) & l\cdot f_2 & l\cdot f_3 & \cdots & \\ & 2l\cdot f_0 & -2l\cdot(1-f_1) & 2l\cdot f_2 & 2l\cdot f_3 & \cdots \\ & & \ddots & \ddots & \ddots & \ddots \\ & & & & & \\ & & & & & \end{pmatrix}。$$

可见X_t是连续时间的时齐的 Markov 链.

6.2　有限格点上的 Ising 模型与 Gauber 动力学

定义６.２９（Ising 模型）

Ising 模型是统计力学中的典型模型。 记$\mathrm{T}=\{1,2,\cdots,N\}$，则$G\overset{\Delta}{=}\mathrm{T}^d$就是每边为$N$个格点的$d$维格点阵列. 设每个格点上有一个粒子，它可以处于$+1$与$-1$两种状态之一，称为两种自旋状态。那么$\mathrm{T}^d$个粒子组成的粒子系统带上各自的一种自旋，就组成粒子系统的一个状态． 全体状态组成的集合记为S，它是一个有限集合． 即

$$S=\{-1,+1\}^G\ =\{h:h=h(x),x\in\mathrm{T}^d,h(x)=-1\text{或}+1\}\,.$$

S中的一个元素是粒子系统的一个状态，在统计力学中有一个专用的名称，称为一个组态。

在组态空间S上定义了如下的一个能量函数：

$$H(h)=-\frac{1}{2}\sum_{x,y\text{相邻}}h(x)h(y)-h\sum_x h(x)\,,$$

其中第一个和为所有相邻格点对上的自旋，对于能量的贡献， 而第二个和表示外场.

令h^x为只在格点x处改变组态h的自旋符号后所得的组态，即

$$h^x(y)=\begin{cases} h(y), & y\neq x \\ -h(x), & y=x \end{cases}.$$

定义６．３０　（Ising 模型的 Glauber 动力学）

Ising 模型的 Glauber 动力学是指如下的转移机制： 当$x\neq h$ 时，只当组态x与组态h间在一个格点上的自旋不同时，才容许x到h的转移， 且这时的转移速率为

$$q_{xh}=\begin{cases} C(x,x), & (h=x^x) \\ 0 & (\text{其它情形}) \end{cases},$$

其中$C(x,x)$可有两种取法：

$$C(x,x)=e^{-b(H(x^x)-H(x))}=e^{-b(\sum_{x,y\text{相邻}}h(x)h(y)+h\sum_x h(x))}$$

或

$$C(x,x)=\frac{1}{1+e^{-b(H(x^x)-H(x))}}.$$

参数b是绝对温度的倒数， 称为倒温度．而 $q_{xx}\overset{\Delta}{=}-\sum_{\text{一切}x}C(x,x)$．因为状态（组态）数是有限的 （共$2^{N^d}$个），Glauber 动力学就是由转移速率矩阵$\mathbf{Q}=(q_{xh})_{x,h\in S}$所决定的连续时间的 Markov 链，其转移矩阵为

$$\mathbf{P}(t)=e^{\mathbf{Q}t}=\mathbf{I}+\frac{\mathbf{Q}}{1!}\mathbf{t}+\frac{\mathbf{Q}^2}{2!}\mathbf{t}^2+\cdots.$$

下面求 Ising 模型的 Glauber 动力学的可逆分布．令

$$a(x)\overset{\Delta}{=}e^{-bH(x)}.$$

它是$\mathbf{Q}$配称列，即 $a(x)q_{xh}=a(h)q_{hx}$． 所以, Glauber 动力学的转移速率矩阵$\mathbf{Q}$确定的 Markov 链具有可逆分布$p=(p_x:x\in S)$，而S为状态空间（全体组态），其中

$$p_x=\frac{e^{-bH(x)}}{\sum_{h\in S}e^{-bH(h)}}.$$

p 称为此 Ising 模型的 Gibbs 分布．

进而，由于转移速率矩阵$\mathbf{Q}$的互通性，运用定理 6.２６得到转移概率矩阵有极限$\mathbf{P}(t)\xrightarrow{t\to\infty}=\mathbf{1}^Tp$．可见 Gibbs 分布是 Glauber 动力学的极限分布．

6．3　生灭类过程

定义６．３１　状态空间为整数，且转移速率矩阵为

$$\mathbf{Q}=\begin{pmatrix} \ddots & \ddots & & \ddots & & & \\ & m_i & -(l_i+m_i) & l_i & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & m_j & -(l_j+m_j) & l_j & \\ & & & & \ddots & \ddots & \end{pmatrix}$$

的连续时间的 Markov 链，称为生灭过程，其中 $l_i, m_i > 0$，$\{l_i + m_i\}$为有界序列．i 为整数．

此时 $\mathbf{Q}$ 是互通的，有配称列 $a=(\cdots,a_i,\cdots,a_j,\cdots)$，满足 $a_i l_i = a_{i+1} m_{i+1}$．即可取

$$a_0=1, a_{i+1}=a_i\frac{l_i}{m_{i+1}},(i\ge 0),\ a_j=a_{j+1}\frac{m_{j+1}}{l_j},(j<0).$$

如果还满足

$$Z\overset{\Delta}{=}1+\sum_{i=0}^{\infty}\prod_{k=0}^{i}\frac{l_k}{m_{k+1}}+\sum_{j=-\infty}^{-1}\prod_{k=j}^{-1}\frac{m_{k+1}}{l_k}<\infty,$$

那么 $\mathrm{p}=\frac{1}{Z}a$ 就是唯一的可逆分布．此时生灭过程的转移矩阵有 $\mathbf{P}(t)\xrightarrow{t\to\infty}\mathbf{1}^T a$．

例６．３２　状态空间为非负整数，且转移速率矩阵为

$$\mathbf{Q}=\begin{pmatrix} -l_0 & l_0 & & & & \\ m_1 & -(l_1+m_1) & l_1 & & & \\ & m_2 & -(l_2+m_2) & l_2 & & \\ & & \ddots & \ddots & \ddots & \\ & & & m_i & -(l_i+m_i) & l_i \\ & & & & \ddots & \ddots \end{pmatrix}.$$

的连续时间的 Markov 链，称为单侧生灭过程，其中 l_0，$l_i, m_i > 0$，$\{l_i + m_i\}$为有界序列，$i \ge 1$．类似地，可以得到存在可逆分布的条件．

有一类特例，就是空间齐次情形，即 $l_i = l_0 = l,, m_i = m$，$(\forall i \ge 1)$ 的情形．此时存在可逆分布的充要条件简化为 $l < m$．在条件成立时，可逆分布为

$$\mathrm{p}=\frac{1}{1-\frac{l}{m}}(1,\frac{l}{m},\cdots,(\frac{l}{m})^n,\cdots).$$

例６．３３　（线性纯生过程，Yule 过程）

考察一种初等生物群体的繁殖过程模型．假定每个生物体与其它生物体的繁殖相互独立地服从参数为 l 的 Poisson 过程，且每个生物体分别地繁殖后代，且不会死亡(可以理解为转变为另一种形式保存下来)．记 N_t 为时刻 t 时该生物群体的个数，称为 Yule 过程．令

$$p_n(t) = P(N_t = n) \qquad n = 0,1,2,\cdots$$

在 $N_t = n$（即已知在 t 时刻有 n 个生物体）的条件下，我们来考察微小的时间段 $(t, t+h]$ 内群体个数的变化规律。由 Poisson 过程的直观含义可知，此过程相当于 n 个相互独立 Poisson 过程的叠加，根据第 3 章的结论，它服从参数 $n\ l$ 为的 Poisson 过程. 于是，在这段时间内繁殖两个或两个以上后代的概率为 $o(h)$，且在 $t+h$ 时的生物体的个数满足

$$\begin{aligned} p_n(t+h) = P\{N_t = n,\ (t, t+h]\text{内无后代出生}\} + \\ + P\{N_t = n-1, (t, t+h]\text{内出生的后代个数为}1\} + o(h) \end{aligned}$$

$$= p_n(t)(1 - n\,l \cdot h) + p_{n-1}(t)(n-1)\,l \cdot h + o(h)$$

在上式两边分别减去 $p_n(t)$，除以 h，令 $h \to 0$，我们得到

$$\frac{d}{dt} p_n(t) = -n\,l p_n(t) + (n-1)\,l p_{n-1}(t), \qquad \text{(Y.1)}$$

其初始条件由时刻 0 的时生物群体的个数所决定. 假定 $N_0 = n_0$，则方程(Y.1)的初始条件为

$$p_{n_0}(0) = 1; \quad p_n(0) = 0. \quad (n \neq n_0) \qquad \text{(Y.2)}$$

在 $n < n_0$ 时，对于任意的 $t \geq 0$，显见有 $p_n(t) = 0$. 在 $n = n_0$ 时，则根据方程（Y.1）及（Y.2）式知

$$\frac{d}{dt} p_{n_0}(t) = -n_0\,l p_{n_0}(t), \qquad p_{n_0}(0) = 1.$$

此方程的解为

$$p_{n_0}(t) = e^{-n_0 l t}. \qquad \text{(Y.3)}$$

而在 $n \geq n_0 + 1$ 时，我们可由方程(Y.1)建立一个递推关系：假定 $p_{n-1}(t)$ 为已知函数，并以

$p_n(0) = 0$ 为初始条件，则可由方程(Y.1)解出

$$p_n(t) = \int_0^t e^{-nl\,(t-s)}(n-1)\,l p_{n-1}(s)ds.$$

最后，由这个递推式及（Y.3）式，我们用归纳法可依次求得

$$p_n(t) = \frac{(n-1)!}{(n-n_0)!(n_0-1)!}(e^{-l t})^{n_0}(1 - e^{-l t})^{n-n_0}, \qquad (n \geq n_0 + 1).$$

以上说明生灭过程 N_t 作为连续时间的 Markov 链具有如下的转移：

$$p_{ij}(t)=\begin{cases} 0 & (j<i) \\ e^{-ilt} & (j=i) \\ \dfrac{(j-1)!}{(j-i)!(i-1)!}(e^{-lt})^{i}(1-e^{-lt})^{j-i} & (j>i) \end{cases}.$$

这时转移速率矩阵为

$$\mathbf{Q}=\begin{pmatrix} 0 & 0 & & & & & \\ & -l & l & & & & \\ & & -2l & 2l & & & \\ & & & \ddots & \ddots & & \\ & & & & -il & il & \\ & & & & & \ddots & \end{pmatrix}.$$

例６.３４　（纯灭过程）

若一个系统由n个部件并联构成,系统的故障率正比于正在工作但未出故障的部件的数目k，即此时的故障率为kl_k。记x_t为在时刻t正在工作而未出故障的部件的数目．令

$$p_k(t)=P(x_t=k).$$

假定在$(t,t+h]$中有两个以上的部件发生故障的概率为$o(h)$， 那么我们有

$$P(x_{t+h}=k-1\,|\,x_t=k)=kl_kh+o(h).$$

用全概率公式，对于$k<n$我们得到

$$p_k(t+h)=(k+1)l_{k+1}h)p_{k+1}(t)+(1-kl_kh)p_k(t)+o(h).$$

两边减去$p_k(t)$后除以h，令$h\to 0$， 便得

$$p_k'(t)=(k+1)l_{k+1}p_{k+1}(t)-kl_kp_k(t)，\quad （k<n）.$$

再则

$$p_n'(t)=-nl_np_n(t).$$

初始条件为

$$p_n(0)=1,\ p_k(0)=0,(k<n).$$

由此我们可以向后归纳地解出这组方程．然而，更简单的是用 Laplace 变换方法．令

$$\hat{p}_k(z)=\int_0^{\infty}p_k(t)e^{-zt}dt.$$

于是由上面的常微分方程组及初值条件立得

$$z\hat{p}_n(z)-1=-nl_n\hat{p}_n(z),$$

$$z\hat{p}_k(z)=(k+1)l_{k+1}\hat{p}_{k+1}(z)-kl_k\hat{p}_k(z).$$

由此解得

$$\hat{p}_n(z)=\frac{1}{z+nl_n}，\quad \hat{p}_k(z)=\frac{n!l_nl_{n-1}\cdots l_{k+1}}{k!(z+nl_n)\cdots(z+kl_k)}.$$

这个例子可以有如下的应用： 由于系统是并联的，所以系统的可靠函数（系统在时刻 t 在工作的概率）为 $1-p_0(t)$ ，由此可以通过查 Laplace 变换表，反过来求得 $p_0(t)$. 为此记

$$a(x)=(x+l_1)(x+l_2)\cdots(x+l_n)，$$

则可得到

$$p_0(t)=1-n!l_1\cdots l_n\sum_{k=1}^{n}\frac{e^{-kl_kt}}{kl_ka'\{-kl_k)}.$$

记系统的寿命为 V ， 它是一个随机变量. 于是系统的平均寿命为

$$m=EV=\int_0^\infty P(V>t)dt=\int_0^\infty(1-p_0(t))dt=\lim_{z\to 0}(\frac{1}{z}-\hat{p}_0(z))=\sum_{k=1}^{n}\frac{1}{kl_k}.$$

*例６．３５　（带迁移的线性生灭过程）

若

$$\mathbf{Q}=\begin{pmatrix}-a & a & & & & \\ m & -(a+l+m) & a+l & & & \\ & 2m & -(a+2l+2m) & a+2l & & \\ & & \ddots & \ddots & \ddots & \\ & & & im_i & -(a+il+im) & a+il \\ & & & & \ddots & \ddots\end{pmatrix}.$$

其中 $m>0$ 为灭绝率， $l>0$ 为生长率，$a>0$ 为迁移率. 为了使系统有平稳性，我们假定 $l<m$.

首先，$\mathbf{Q}$ 有配称列 $m=(m_0,m_1,\cdots)$ ，它满足：

$$m_0a=m_1m$$
$$\cdots\cdots$$
$$m_n(a+nl)=m_{n+1}(n+1)m.$$

所以可取

$$m_0 = 1$$
$$m_1 = \frac{a}{m}$$
$$\cdots\cdots$$
$$m_n = m_{n-1}\frac{a+(n-1)l}{nm} = \cdots = \frac{[a+(n-1)l]\cdots[a+l]a}{n!m^n}$$
.

于是

$$Z \stackrel{\Delta}{=} 1+\sum_{n=0}^{\infty}\frac{[a+(n-1)l]\cdots(a+l)a}{n!m^n} < \infty .$$

从而 $p = \frac{1}{Z}m$ **是一个分布.**

再则，对于 $q_n = a + n(l+m)$ **，有**

$$\sum_n q_n m_n < \infty$$

这说明定理 6.28 条件满足．因此 $\mathbf{Q}$ 是一个对称的连续时间的 Markov 链的转移矩阵 $P(t)$ **的速率矩阵**.

我们将对固定的 i **用向前方程求出** $\{p_{ij}(t), j\in S\}$ **的母函数**

$$P_i(t,z) \stackrel{\Delta}{=} \sum_{j=0}^{\infty} p_{ij}(t)z^j .$$

注意向前方程' P '(t)=P(t)Q ' 的分量形式为 $(j \ge 1)$

$$p_{i,0}'(t) = -ap_{i,0}(t) + p_{i.1}(t)$$

$$p_{ij}'(t) = -(a + j(l+m))p_{ij}(t) + (a+(j-1)l)p_{i,j-1}(t) + (j+1)mp_{i,j+1}(t) ,$$

其初值为 $p_{ij}(0) = d_{ij}$ **. 于是** $P_i(t,z)$ **满足**

$$\begin{cases} \dfrac{\partial P_i}{\partial t} = a(z-1)P_i + (m-lz)(1-z)\dfrac{\partial P_i}{\partial z} \\ P_i(0,z) = z^i \end{cases} .$$

这是一个一阶线性偏微分程,可以通过特征线法求解．为此，先把方程改写成

$$\frac{\partial P_i}{\partial z} + \frac{-1}{(m-lz)(1-z)}\frac{\partial P_i}{\partial t} = \frac{a}{m-lz}P_i .$$

特征线法的思想是引进一条(实际上是一族)曲线 *t=t(z)*，使 $P_i(t,z)$ **在 *t(z)*上可以表成** z **的函数．我们**

想要选 *t(z)*，使方程左方成为 $\frac{dP_i(t(z),z)}{dt}$ **(即** $\frac{\partial P_i}{\partial t}t'(z) + \frac{\partial P_i}{\partial z}$ **)．为此只需选**

$$t'(z)=\frac{-1}{(m-lz)(1-z)}=\begin{cases}\dfrac{-1}{l(1-z)^2} & (m=l)\\ \dfrac{1}{m-l}(\dfrac{1}{1-z}+\dfrac{l}{m-lz}) & (m\neq l)\end{cases}.$$

(注意 $m>l>0$)．从此解出 *t(z)*得：

$$t(z)=C+\begin{cases}\dfrac{1}{l(z-1)} & (m=l)\\ \dfrac{1}{l-m}\log|\dfrac{m-lz}{1-z}| & (m\neq l)\end{cases}$$

$$\overset{\Delta}{=}C+t_0(z).$$

这样，$P_i(t(z),z)$ 满足方程

$$\frac{dP_i(t(z),z)}{dt}(=\frac{\partial P_i}{\partial z}+\frac{-1}{(m-lz)(1-z)}\frac{\partial P_i}{\partial t})=\frac{a}{m-lz}P_i.$$

此方程是可分离系数的，它的解为

$$P_i(t(z),z)=K|m-lz|^{-\frac{a}{l}}.$$

由于这个解是沿依积分常数 C 的曲线族 t(z)求的，所以这个任意常数 K 必须依赖于 C，即应该有

$$K=K(C)=K(t(z)-t_0(z)).$$

于是

$$P_i(t,z)=K(t-t_0(z))|m-lz|^{-\frac{a}{l}}.$$

为了确定函数 K 的具体形式，需要利用初值 $P_i(0,z)=z^i$，此即

$$K(-t_0(z))=z^i|m-lz|^{\frac{a}{l}}.$$

再作变换 $u=-t_0(z)$，便得

$$K(u)=(t_0^{-1}(-u))^i|m-lt_0^{-1}(-u)|^{\frac{a}{l}}.$$

把它代入 $P_i(t,z)$ 的表达式，最后便得到 $P_i(t,z)$ 的表达式．

由定理 6.28，我们有

$$\mathbf{P}(t)\xrightarrow{t\to\infty}\mathbf{1}^T p\ .$$

如果 $a=0$，即无迁移，那么这是 0 点为吸收点的空间齐次的单侧生灭过程．

[注] 若条件 $m>l$ 不满足，则易见配称列各分量之和为 ∞，不可能满足定理 6.28 的条件条件 ．但是，可

以与前面类似地仍可求得 $P_i(t,z)$ 的明显表达式. 并求得此 Markov 链的灭绝概率.

6.4 系统与有效度

一个系统常由一些串联, 并联, 桥联或带反馈的各个部件构成. 每个部件独立地工作, 且近似地假定其应用寿命为服从指数分布的随机变量(如果各部件的寿命的分布不是指数分布，则本段的方法不适用). 系统在时刻t还在工作的概率就称为它的可靠性函数. 典型的应用出于设备维修.

例6.36 (典型例子1 — 作为双侧生灭过程的设备系统)

有 M 台设备独立地工作,每台设备的正常运转时间段为服从参数λ的指数随机变量, 且相互独立. 当运转不正常时就需要检修. 假定有 $n(n\le M)$ 个修理工. 记 N_t 为在时刻 t 此系统中不能正常运转的设备数(包括正在修理的与待修的). 假定修理工相互独立地, 以服从参数μ的随机时间修好系统中的一个设备. 我们需要求 N_t 的不变分布, 它描述了系统在稳定时的工作质量, 由此可以对于备用件的储备方案提供指导意见.

解 (1)设备并联的情形.

N_t 的可能取值为 $0,1,\cdots,M$. 由指数流的性质我们有: 每台设备在 $(t,t+h]$ 中需检修 1 次的概率等于 $l\cdot h+o(h)$, 而需检修 2 次以上的概率为 $o(h)$, 又每台在修设备在 $(t,t+h]$ 中已修好的概率为 $m\cdot h+o(h)$, 未修好的概率为 $1-m\cdot h+o(h)$. 利用 k 个独立同指数分布(参数为 r)随机变量的最小值服从参数为 kr 的指数分布这一事实, 考虑到设备运转与检修时间之间的独立性, 就可求得

$$P(N_{t+h}=j\mid N_t=i)=\begin{cases}(M-i)\,lh+o(h) & (j=i+1)\\ (n\wedge i)mh+o(h) & (j=i-1)\\ o(h) & (|j-i|>1)\end{cases}.$$

于是 N_t 是具有下面的转移速率矩阵的连续时间的 Markov 链:

$$\mathbf{Q}=\begin{pmatrix}-Ml & Ml & & & & \\ \ddots & \ddots & & \ddots & & \\ & (n\wedge i)m & -[(M-i)l+(n\wedge i)m] & (M-i)l & & \\ & & \ddots & & \ddots & \ddots \\ & & & & nm & -nm\end{pmatrix},$$

$\mathbf{Q}$ 是互通的, 且具有配称列 $m=(m_0,\cdots,m_M)$, 它满足

$$m_i(M-i)l=m_{i+1}(n\wedge(i+1))m.$$

由此得到

$$m_{i+1}=\frac{M-i}{n\wedge(i+1)}\frac{l}{m}m_i=\cdots=\left(\frac{l}{m}\right)^{i+1}\prod_{k=0}^{i}\frac{M-k}{n\wedge(k+1)}m_0.$$

不妨假定 $m_0=1$**．令**

$$Z\overset{\Delta}{=}1+\sum_{i=1}^{M}(\frac{l}{m})^i\prod_{k=0}^{i-1}\frac{M-k}{n\wedge(k+1)},\quad p\quad=\frac{1}{Z}\ m\ .$$

那么

$$p_i=\begin{cases}C_M^i(\frac{l}{m})^ip_0 & (i\le n)\\ C_M^i(\frac{l}{m})^i\frac{(n+1)(n+2)\cdots i}{n^{i-n}} & (n<i\le M)\end{cases}.$$

如果 $n=M$ **，那么由上面可推出**

$$p_0=(1+\frac{l}{m})^{-M}=(\frac{m}{m+l})^M$$

$$p_i=C_M^i\left(\frac{l}{m}\right)^i\left(\frac{m}{m+l}\right)^M=C_M^i\left(\frac{l}{m+l}\right)^i\left(\frac{m}{m+l}\right)^{M-i}.$$

这就是说，作为极限分布的可逆分布 p **，是二项分布** $B(n,p)$ **， 其中** $n=M$ **为独立试验的次数，** $p=\dfrac{l}{m+l}$ **为在每次试验中所涉事件发生的概率．这样** $p=\dfrac{l}{m+l}$ **可以解释为每台设备待修的概率，而** $\dfrac{m}{m+l}$ **是每台设备不需检修的概率 — 即单台设备的有效度．这说明在并联情形，当** $n=M$ **且稳定时，即使有修理工的介入，各台设备的有效度是相互独立的．**

由定义可知 p **是此连续时间的 Markov 过程的可逆分布．而且** $\mathbf{P}(t)\xrightarrow{t\to\infty}\mathbf{1}^Tp$ **．特别地对于任意状态** i **，有极限关系**

$$p_{i,M}(t)\xrightarrow{t\to\infty}p_M=\frac{1}{Z}\left(\frac{l}{m+l}\right)^M,$$

它是在 $n=M$ **且时间充分长时，系统故障的概率．**

在 $n=M$ **的情形，我们不但可以算出** N_t **的不变分布(也称为平稳分布)** p **，还可以算出在** $N_0=0$ **的条件下** N_t **的条件分布：**

$$\boldsymbol{p}(t)\overset{\Delta}{=}(p_0(t),\cdots,p_M(t)),\quad p_k(t)=P(N_t=k\mid N_0=0).$$

事实上,只要利用母函数方法去求向前方程 P'(t)=P(t)Q 的第一行的解，经过不太繁琐的计算便可得到 $\boldsymbol{p}(t)$ **是参数** $n=M,p=\dfrac{l}{m+l}(1-e^{-(l+m)t})$ **的二项分布** $B(n,p)$ **．从而可以**

定义一台设备在时刻t的瞬时有效度为$1-p$，即 1-$\frac{l}{m+l}(1-e^{-(l+m)t})$.

(2)设备组成一个串联系统.

令状态0与1分别表示系统处在正常工作与处在检修状态.记x_t为系统在时刻t所处的状态.于是x_t可取值的状态集为$S=\{0,1\}$，且具有转移速率矩阵

$$\mathbf{Q}=\begin{pmatrix}-Ml & Ml\\ m & -m\end{pmatrix}.$$

它是互通的，并有可逆分布

$$p=\frac{1}{1+\frac{Ml}{m}}(1,\frac{Ml}{m}).$$

在稳定情形下系统正常运转的概率即为p_0，所以这时系统的有效度为 $p_0=\frac{m}{m+Ml}$.

注意$\frac{m}{m+Ml}>(\frac{m}{m+l})^M$，后者正好是各台设备独立情形的有效度.可见由于修理工的修理打破了独立性，从而增加了有效度.再则，在串联情形的计算中，修理工的数目n并未出现，这说明在实际上只需一个修理工就够了，显见其它$n-1$个是多余的.

如果我们知道了停产的损失费用及设置一个修理工的工资，那么,通过本例的分析，可以帮助我们设计在稳定情形下修理工的最佳配置数目.

例6.37 （典型例子2 — 含不同型设备且修理工数与设备数相等的系统）

n台设备分别独立地运转.设第k台设备的正常运转时间按参数l_k的指数分布,而当它不正常时需经参数为m_k的指数分布随机时间修理才能恢复正常运转.设有n个修理工，且修理所需的时间与正常运转时间之间独立.求系统在稳定情形时的有效度.

解　先设$n=2$.

(1) 系统中各设备并联的情形.

取状态空间 S=｛0，1，2，3｝，其中“0”表示两台设备都正常，″1″，“2″分别表示第 1，2 台设备不正常，“3” 表示两台设备都不正常.令N_t在 S 的取值表示系统在时刻t所处的状态.

用与前面类似的推导可得N_t是具有以下转移速率矩阵的连续时间的 Markov 链：

$$\mathbf{Q}=\begin{pmatrix} -(l_1+l_2) & l_1 & l_2 & 0 \\ m_1 & -(m_1+l_2) & 0 & l_2 \\ m_2 & 0 & -(m_2+l_1) & l_1 \\ 0 & m_2 & m_1 & -(m_1+m_2) \end{pmatrix}.$$

Q 是互通的，具有配称列

$$\mathrm{a}=(a_0,a_1,a_2,a_2)\text{，}$$

它满足

$$a_0 l_1 = a_1 m_1\text{，}a_0 l_2 = a_2 m_2\text{，}a_1 l_2 = a_3 m_2\text{，}a_2 l_1 = a_3 m_1\text{.}$$

即

$$a_1=\frac{l_1}{m_1}a_0\text{，}a_2=\frac{l_2}{m_2}a_0\text{，}a_3=\frac{l_1}{m_1}\frac{l_2}{m_2}a_0\text{.}$$

令

$$\boldsymbol{\pi}=(p_0,p_1,p_2,p_3)=(1+\frac{l_1}{m_1}+\frac{l_2}{m_2}+\frac{l_1 l_2}{m_1 m_2})^{-1}(1,\frac{l_1}{m_1},\frac{l_2}{m_2},\frac{l_1 l_2}{m_1 m_2})\text{，}$$

它就是 N_t 的可逆分布. 从而稳定情形时两台设备并联系统的有效度为

$$1-p_3=1-\frac{l_1}{m_1+l_1}\frac{l_2}{m_2+l_2}\text{.}$$

这也说明并联情形在稳定时可以看成各台设备是独立运转的.

(2) **系统中各设备串联的情形**

此时取状态空间S=｛0，1，2｝，其中“0”表示两台设备都正常，″1″，“2″分别表示第 1，2 台设备不正常. 令 N_t 在 S 的取值表示系统在时刻 t 所处的状态. 用与并联情形类似的推导可得，N_t 是具有以下转移速率阵的 Q 过程：

$$\mathbf{Q}=\begin{pmatrix} -(l_1+l_2) & l_1 & l_2 \\ m_1 & -m_1 & 0 \\ m_2 & 0 & -m_2 \end{pmatrix}.$$

Q 是互通的，具有配称列

$$\mathrm{a}=(a_0,a_1,a_2)\text{，}$$

它满足

$$a_0 l_1 = a_1 m_1\text{，}a_0 l_2 = a_2 m_2\text{.}$$

即

$$a_1 = \frac{l_1}{m_1} a_0 \text{，} a_2 = \frac{l_2}{m_2} a_0 \text{.}$$

令

$$\boldsymbol{\pi} = (p_0, p_1, p_2) = (1 + \frac{l_1}{m_1} + \frac{l_2}{m_2})^{-1} (1, \frac{l_1}{m_1}, \frac{l_2}{m_2}),$$

它就是 N_t 的可逆分布. 而稳定情形时两台设备串联系统的有效度为 p_0, 即在稳定情形时的串联有效度为 $(1 + \frac{l_1}{m_1} + \frac{l_2}{m_2})^{-1}$. 注意 $(1 + \frac{l_1}{m_1} + \frac{l_2}{m_2})^{-1} > \frac{m_1}{m_1 + l_1} \frac{m_2}{m_2 + l_2}$. 可见这时比设备独立运转时提高了有效度. 此时实际上并不需要两个修理工, 而只需一个.

对一般有 $n > 2$ 的情形，可以类似地得到相应的结论．令

0 表示系统正常工作, k 表示第 k 台设备在修理，$(0 < k \le n)$，

N_t 为系统在时刻 t 所处的状态，它取值于状态空间 $S = \{0,1,\cdots,n\}$．与前面类似地可以证明 N_t 是具有下述转移速率矩阵的连续时间的 Markov 链：

$$\mathbf{Q} = \begin{pmatrix} -(l_1 + \cdots + l_n) & l_1 & \cdots & l_n \\ m_1 & -m_1 & & \\ \vdots & & \ddots & \\ m_n & & & -m_n \end{pmatrix}.$$

Q 也是互通的，具有配称列

$$\mathrm{a}, = (a_0, a_1, \cdots, a_n),$$

它满足

$$l_k a_0 = m_k a_k, \quad (1 \le k \le n).$$

于是

$$\boldsymbol{\pi} \overset{\Delta}{=} (1 + \frac{l_1}{m_1} + \cdots + \frac{l_n}{m_n})^{-1} \mathrm{a}$$

就是 N_t 的可逆分布．所以系统在稳定情形时的有效度为：

$$p_0 = (1 + \frac{l_1}{m_1} + \cdots + \frac{l_n}{m_n})^{-1}.$$

显见，由于上式右方大于各设备独立运转时的有效度 $\prod_{k=1}^{n} (1 + \frac{l_k}{m_k})^{-1}$.

例6．38　(具有备用设备的系统)

设一个冷贮系统有 n 台相同设备组成，其中一台在工作，其它 $n-1$ 台均为备用设备.车间设有一条修理线.假定设备正常运转时间服从参数为λ的指数分布，而修理设备的时间服从参数为μ的指数分布，它们之间相互独立．求系统在稳定时的有效度.

解 令 N_t 为在时刻 t 在修与待修的设备台数之和，其状态空间为 $S=\{0,1,\cdots,n\}$．仿前面的推导可知 N_t 是具有下述转移速率矩阵的连续时间的 Markov 链:

$$\mathbf{Q}=\begin{pmatrix} -l & l & & & \\ m & -(l+m) & l & & \\ & \ddots & \ddots & \ddots & \\ & & m & -(l+m) & l \\ & & & m & -m \end{pmatrix}.$$

$\mathbf{Q}$ 是互通的，具有可逆分布

$$\boldsymbol{\pi}=(p_0,p_1,\cdots,p_n),$$

其中

$$p_0=(1+\frac{l}{m}+\cdots+(\frac{l}{m})^n)^{-1},p_k=(\frac{l}{m})^k p_0,(1\le k\le n).$$

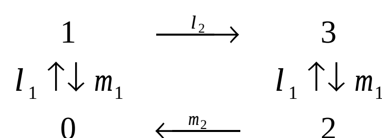
p_0 即是系统在稳定时的有效度.

例6．39（具有性能较差的备用设备的系统）

设冷贮系统由两台设备组成，其中一台是工作设备，另一台是比前一台性能较差的备用设备.车间设有一个修理工。当失效的工作设备被修好时，则立刻用它来代替正在工作的备用设备． 如果工作设备失效，且备用设备也正在修理时，则立刻停修后者，并先修前者.设工作设备与备用设备的正常运转时间分别服从参数为 l_1,l_2 的指数分布。而修理好设备的时间分别服从参数为 m_1,m_2 的指数分布．假定它们之间都是相互独立的。求系统在稳定时的有效度.

解　取状态空间 S=｛0，1，2，3｝，其中“0”表示两台设备都正常，”1”，“2”分别表示第 1，2 台设备不正常，“3”表示两台设备都不正常．令 N_t 在 S 的取值表示系统在时刻 t 所处的状态.

根据优先修理优先使用工作设备的要求，我们画一张示意"流向图"如下:



类似的推导可得到 N_t 是具有下述转移速率矩阵的连续时间的 Markov 链:

$$\mathbf{Q}=\begin{pmatrix} -l_1 & l_1 & & \\ m_1 & -(m_1+l_2) & & l_2 \\ m_2 & & -(m_2+l_1) & l_1 \\ & & m_1 & -m_1 \end{pmatrix}.$$

Q 是互通的，但没有配称列．我们可以通过解联立方程π Q =0，$\mathbf{1}^T p=1$，来得到 N_t 的不变分布π，即

$$\begin{cases} -l_1 p_0+m_1 p_1+m_2 p_2=0 \\ l_1 p_0-(m_1+l_2)p_1=0 \\ -(m_2+l_1)p_2+m_1 p_3=0 \\ p_0+p_1+p_2+p_3=1 \end{cases}.$$

解此方程组我们得到

$$\begin{cases} p_1=\dfrac{l_1}{l_2+m_1}p_0 \\ p_2=\dfrac{l_1}{m_2}\dfrac{l_2}{l_2+m_1}p_0 \\ p_3=\dfrac{l_1 l_2}{m_1 m_2}\dfrac{l_1+m_2}{l_2+p_1}p_0 \end{cases}$$

于是系统在稳定时的有效度为 $1-p_3$．

例 6．40（系统有处于其它状态的同型备用设备）

设热贮系统由两台相同的设备,其中一台处于工作态,另一台处于热贮态。 它们的正常运转时间分别服从参数为λ，υ的指数分布,而设备在不正常时所需修理时间均为参数为μ的指数分布．它们都是独立的． 修理原则为：先失效的设备先修．但是如果工作态设备失效而热贮态设备正常,则后者自动转变为工作态,而且正常运转时间的指数分布的参数由υ自动转变为λ．此后，如果原来状态的设备修复，则此修复的设备又自动转变为热贮状态，并且其运转时间的指数分布的参数自动由λ转变为υ．求这个系统在稳定时的有效度．

解　令状态 0,1,2 分别的含义为： 0 表示两台都正常，1 表示 一台处于工作态，另一台在修理，2 表示两台都不正常(系统失效)． 状态空间 S={1,2,3}，N_t 为系统在时刻 t 所处的状态．

用类似的思路可以推出 N_t 是具有下述转移速率矩阵的联系实践的 Markov 链：

$$\mathbf{Q}=\begin{pmatrix} -(l+v) & l+v & \\ m & -(l+m) & l \\ & m & -m \\ & & \end{pmatrix}.$$

Q 是互通的，具有可逆分布：

$$\boldsymbol{\pi}=(p_0,p_1,p_2),$$

其中

$$p_1=\frac{l+v}{m}p_0, p_2=\frac{l(l+v)}{m^2}p_0, p_0=(1+\frac{l+v}{m}+\frac{l(l+v)}{m^2})^{-1}.$$

于是在系统稳定时的有效度为 $1-p_2$.

7. 连续时间的 Markov 链的模拟与加速收敛

7. 1 连续时间的 Markov 链的轨道的模拟

定理 6. １２提供了连续时间的 Markov 链的轨道的随机模拟方法．我们以只有 N 个状态的连续时间的 Markov 链为例，其轨道可以由如下步骤模拟得到:

(1). 模拟初开始分布 m_0，对于 $1\le i\le N$，以概率 $m_0(i)$ 取 i，其中 $m_0=(m_0(1),\cdots,m_0(N))$；

(2). 以 $\exp_{q_i}$ 分布的时间呆在 i. 然后以概率 $\frac{q_{ij}}{q_i}$ 跳到 $j(\neq i)$；

(3). 以 j 代替 i 后重复(2).

7. 2 加速收敛的均匀化方法

对转移速率矩阵 Q，只要其行和为 0, 且 $q_i\le l \stackrel{\Delta}{=} \sup_j q_j$，我们要找一个快速计算转移矩阵 $\mathbf{P}(t)$ 的表达式。为此我们定义 $\tilde{P}\stackrel{\Delta}{=}\frac{1}{l}Q+I$，易见它是随机矩阵（意即其元素非负，且行和为 1），故对于任意 n，$(\tilde{\mathbf{P}})^n$ 的所有分量都小于 1, 且满足 $Q=l(\tilde{P}-I)$. 于是

$$\mathbf{P}(t)=e^{Qt}=e^{-lt(I-\tilde{P})}=\sum_{\mathbf{n}=0}^{\infty}\tilde{\mathbf{P}}^n\frac{(l\cdot t)^n}{n!}e^{-l\cdot t}$$

是指数收敛的, 且其收敛速度不低于强度为 l 的 Poisson 过程.

习题 6

1. 取非负整值的连续时间的 Markov 链 N_t，如果存在非负函数 $l(t,x)$ 使

$$E[(N_{t+h}-N_t)\,|\,N_t]=l(t,N_t)h+o(h),$$

则称为以 $l(t,N_t)$ 为随机强度的点过程. 证明其随机强度是常数的充要条件为它是 Poisson 过程.

2 若 x_t 为 Yule 过程，满足 $x_0=1$. 设 S_k 为其首达 k 的时刻: $S_k=\inf\{t:x_t=k\}$，$t_k\stackrel{\Delta}{=}S_k-S_{k-1}, t_0=0$. (1) 求 $(t_1,\cdots,t_n)$ 的分布.

(2) 求证在$x_t = n+1$的条件下，$(S_1, \cdots, S_n)$ 有正比于

$$e^{-ls_1} e^{-2l(s_2-s_1)} \cdots e^{-nl(s_n-s_{n-1})} e^{-(n+1)l(t-s_n)} I_{0<s_1<\cdots<s_n<t}$$

的密度. 进一步再把此密度具体地写出来． (提示 仿照 Poisson 过程).

3. 对于参数为 $l_n = nl$，$m_n = nm$ 的生灭过程x_t，记$m_t = Ex_t$，并设 $x_0 = m_0$. 用 Master 方程求出 m_t 满足的方程，并求其解.

4. 设有 $l_n = l$，$m_n = nm$ 的生灭过程x_t，满足$x_0 = 0$. 求证它的矩母函数为

$$Ee^{sx_t} = e^{-\frac{l}{m}(1-e^{-mt})(s-1)}.$$

并由此得到

$$Ex_t = \frac{l}{m}(1-e^{-mt}).$$

5．若独立增量过程N_t取整数值，$N_0 = 0$，且当$h \to 0$时有

$$P(N_{t+h} - N_t = -1) = mh + o(h)\text{，}P(N_{t+h} - N_t = 1) = l.h + o(h)\text{，}$$

$$P(N_{t+h} - N_t = 0) = 1-(m+l)h + o(h).$$

(1)．求它的矩母函数$G(t,z) = \sum_{n=-\infty}^{\infty} P(N_t = n)z^n$满足的偏微分方程，证明

$$G(t,z) = e^{(lz+mz^{-1}-(l+m))t}.$$

再求 $EN_t, Var(N_t)$．并推广这个结论至l, m都依赖于t的情形．

(2)．如果再假定 N_t只取非负整数，且$P(N_h - N_0 = -1) = o(h)$，则N_t是单侧生灭过程，它可以描述某种种群的大小．设初始种群大小为$N_0 = i$．种群灭绝的时刻T就是N_t首达 0 的时刻，它是一个随机变量．证明在$m = l$时的灭绝分布为 $P(T \le t) = \left(\frac{l \cdot t}{l \cdot t + 1}\right)^i$．

6．若独立增量过程N_t取非负整数值，$N_0 = 0$，且当 h→0 时有

$$P(N_{t+h} - N_t = 1) = l_1 h + o(h),$$

$$P(N_{t+h} - N_t = 2) = l._2 h + o(h), P(N_{t+h} - N_t > 2) = o(h).$$

求N_t的矩母函数．再推广这个结论．

7．若独立增量过程 N_t 取非负整数值，$N_0=0$，且当 h→0 时有

$$P(N_{t+h}-N_t=1\,|\,N_t=k)=l_k h+o(h),\ P(N_{t+h}-N_t\geq 2\,|\,N_t=k)=o(h).$$

求 $P_j(t)\overset{\Delta}{=}P(N_t=j),(j\geq 0)$ 满足的微分方程组，再求解.

*8．设非负整值的时齐的独立增量过程 N_t，满足 $N_0=0$，且对于任意 t 有 $0<P(N_t=0)<1$．

又若存在 $d>0$， 及一个不全为 0 的非负可和数列（即满足条件 $\sum_{n-1}^{\infty} a_n<\infty\}$ 使

$$|\,P(N_{t+h}-N_t=n\,|\,N_{t+h}>N_t)-a_n h\,|\leq Ch^{1+d}.$$

那么 N_t 的矩母函数有简单表示式 $G(t,z)=e^{lt(g(z)-1)}$，其中 $l=\sum_{n=1}^{\infty} a_n$，$g(z)=\sum_{n=1}^{\infty} p_n z^n$，

$p_n=\dfrac{a_n}{l}$．并说明 N_t 是取非负整值的复合 Poisson 过程.

9．有 m 台设备，每台的正常运转时间都服从参数为 λ 的指数分布，检修所需时间均服从参数为 μ 的指数分布且与运转情形独立，各台设备也独立．假定采用如下的检修方案： 有 k 台不能正常运转就全部停工检修．记时刻 t 不能正常运转的设备数为 N_t．

(1)．求证 N_t 为连续时间的 Markov 链，其转移速率矩阵为

$$\mathbf{Q}=\begin{pmatrix} -ml & ml & 0 & & \\ 0 & -(m-1)l & (m-1)l & \ddots & \\ & \ddots & \ddots & \ddots & \\ & & 0 & -(m-k+1)l & (m-k+1)l \\ m & & & 0 & -m \end{pmatrix}.$$

(2)．求 N_t 的平稳分布.

10．设 $\mathbf{P}(t)$ 是只有有限个状态连续时间的 Markov 链的转移矩阵，其转移速率矩阵为 $\mathbf{Q}$．证明

(1) 它的行列式是正的.

(2) 0 是 $\mathbf{Q}$ 特征值，而且 $\mathbf{Q}$ 的非 0 特征值的实部为负.

(3)如果 $\mathbf{Q}$ 对称，且 $\mathbf{P}(t)$ 的所有分量全正，那么对于任意 i，对于如下定义的熵函数

$H_i(t)=-\sum_j p_{ij}(t)\log p_{ij}(t)$ 有

(a) $p_i{}'(t)=\sum_{j=1}^{N} q_{ij}(p_j(t)-p_i(t))$.

(b) $H'(t)=\frac{1}{2}\sum_{i,j=1}^{N} q_{ij}(p_j(t)-p_i(t))(\log p_j(t)-\log p_i(t))\geq 0$.

即 $H(t)$ 是 t 的递增函数.

11. 设 $\mathbf{P}(t)$ 是只有两个状态的转移矩阵，且至少一个不是吸收态. 那么它必定为下述形式

$$\mathbf{P}(t)=\frac{1}{a+b}\begin{pmatrix} a+be^{-(a+b)t} & b(1-e^{-(a+b)t}) \\ a(1-e^{-(a+b)t}) & b+ae^{-(a+b)t} \end{pmatrix}(a,b\geq 0, a+a>0).$$

12. 设带有迁移 a 的取非负整值的线性纯生过程 x_t 的转移速率矩阵为

$$\mathbf{Q}=\begin{pmatrix} -a & a & & & \\ & -(a+l) & (a+l) & & \\ & & -(a+2l) & a+2l & \\ & & & \ddots & \ddots \\ & & & & \end{pmatrix}(a, l>0).$$

证明其转移矩阵 $\mathbf{P}(t)$ 为上三角形矩阵，满足 $p_{ii}(t)=e^{-(a+il)t}$，而当 $i<j$ 时则有

$$p_{ij}(t)=\frac{1}{(j-i)!\,l^j}(a+il)(a+[i+1]l)\cdots(a+[i+j-1]l)e^{-(a+il)t}(1-e^{-lt})^{j-i}.$$

13. 如果在生灭过程 N_t 的转移速率矩阵中的参数为

$l_n=nl+a, m_n=nm \quad (n\geq 0, a, l, m>0)$. 证明

(1) $E(N_t\mid N_0=i)=\begin{cases}\frac{a}{l-m}[e^{(l-m)t}-1]+ie^{(l-m)t} & (l\neq m)\\ i+at & (l=m)\end{cases}$.

(2) 当 $l<m$ 时有平稳分布:

$$p_0=(1-\frac{l}{m})^{\frac{a}{l}},\ p_n=\frac{m^n}{n!}a(a+l)\cdots(a+[n-1]l)(1-\frac{l}{m})^{\frac{a}{l}}.$$

14. 带随机线性自调制 lN_t 的 Poisson 过程，称为二重 Poisson 过程. 证明不带迁移的纯生过程是其特例.

15. 证明取非负整值的连续时间的 Markov 链是时齐的独立增量过程的充要条件为其转移速率矩阵有如下的形式

$$\mathbf{Q}=\begin{pmatrix} -l & l\cdot p_1 & l\cdot p_2 & l\cdot p_3 & \cdots \\ & -l & l\cdot p_1 & l\cdot p_2 & \ddots \\ & & -l & l\cdot p_1 & \ddots \\ & & & \ddots & \ddots \\ & & & & \end{pmatrix} \quad (l,>0, p_i\geq 0, i\geq 1, \sum_{i=1}^{\infty} p_i=1).$$

再求转移矩阵$P(t)$的表达式.

16 如果机器人可处于忙态与闲态，在时间区间$(t,t+h)$转变状态一次的概率为$l\cdot h+o(h)$，而转变两次以上的概率为$o(h)$．假定有n个机器人独立地工作，并且在$t=0$是全处在闲态．证明在时刻t恰有m个机器人处于闲态的概率$p_m(t)$满足方程

$$p'_m(t)=l(n-m+1)p_{m-1}(t)-l\cdot np_m(t)+l(m+1)p_{m+1}(t) \quad (p_{-1}(t)=p_{n+1}(t)=0).$$

记在时刻t处于闲态的机器人数目为N_t．证明 N_t服从二项分布$B(n,\dfrac{1-e^{-lt}}{2})$．

17．如果连续时间的 Markov 链的转移速率矩阵$Q=(q_{ij})$保守(即行和为 1)，而且其对角线上的分量相同，都是$-l$，证明它的转移矩阵有如下的形式:

$$\mathbf{P}(t)=\sum_{n=0}^{\infty}(\tilde{\mathbf{P}})^n\frac{(l\cdot t)^n}{n!}e^{-lt},$$

其中$\tilde{\mathbf{P}}$是此 Markov 链的跳跃链的转移矩阵: $\tilde{p}_{ij}=\dfrac{q_{ij}}{l} \quad (j\neq i), \tilde{p}_{ii}=0$. 解释此公式的概率含义.

18．设某种物种有迁入，死亡或繁衍，每次繁衍只出现一个子代，同时自己就死亡．也在并不繁衍时随机死亡．繁衍，随机死亡，迁入都是相互独立的．假定在时间区间$(t,t+h]$中

(1) 每个个体繁衍的概率为$l\cdot h+o(h)$，此时产生的子代在此时间内繁衍或随机死亡的概率为$o(h)$；

(2) 每个个体随机死亡的概率为$m\cdot h+o(h)$；

(3) 从外面迁入一个个体的概率为$ah+o(h)$，迁入两个以上的概率为$o(h)$ $(l,m,a>0)$．

记在时刻t的个体总数为N_t，记其矩母函数为 $G(t,z)=\sum_{n=0}^{\infty}P(N_t=n)z^n$．证明

(1) 它满足方程

$$\frac{\partial G}{\partial t}=(l\cdot z-m)(z-1)\frac{\partial G}{\partial z}+a(z-1)G.$$

(2) 记 $m(t)=EN_t=\frac{\partial G}{\partial z}(t,1)$，证明 $m(t)$ 满足 $m'(t)=(l-m)m(t)+a$．并求出 $m(t)$．

(3) 问 N_t 是否是点过程，其随机强度是什么?

(4) 求 N_t 的转移速率矩阵．证明它有平稳分布的充要条件为 $m>l$．在条件成立下求平稳分布．

(5) 假定 $l=0$，$N_0=i$．证明(1)中方程仍然成立，并且其解为

$$G(t,z)=(1-(z-1)e^{-mt})^i e^{\frac{a}{m}(z-1)(1-e^{-mt})}$$，而其平稳分布为 $Poisson_{\frac{a}{m}}$。

19．若 $x_t=x_0(-1)^{N_t}$，其中 $\{N_t\}$ 为与 x_0 独立的非时齐的 Poisson 过程，而 $x_0\sim\begin{pmatrix}-1 & 1\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}$．证明

(1) 对于 $s<t$，相关函数 $B(s,t)\overset{\Delta}{=}E(x_sx_t)=e^{-\int_s^t l(u)du}$；

(2) (x_s,x_t) 的特征函数 $Ee^{i(q_1x_s+q_2x_t)}=\cos q_1\cos q_2-B(s,t)\sin q_1\sin q_2$；

如果 N_t 还是时齐的，证明 x_t 是 Markov 链．再求它的转移速率矩阵及不变分布，并证明它是可逆的．

龚光鲁，钱敏平著 应用随机过程教程及其在算法与智能计算中的应用

清华大学出版社,2003

第 7 章　排队过程简介

1．排队过程的描述

1．1　排队系统

排队模型是一种包含更新过程与生灭过程机制的，但是更为复杂的概率模型．

简单的排队过程是在两个相互独立的流作用下形成的，其中一个是要求服务的 "顾客流"，这时假定顾客是一个一个地到达的，其时间间隔组成一个更新流．另一个是当顾客进入服务线后，接受服务的服务时间流． 服务不一定一个接着一个地发生，在两次服务之间可能有空隙,所以虽然各顾客接受的服务独立同分布的时间流．但是它不具有更新流的特点．这些服务空隙成为排队系统的闲期．两个闲期之间的随机时间称为忙期．平均闲期长度与平均忙期长度是排队系统设计中的重要指标.我们如果无视闲期的存在，而虚拟地把服务流一个接着一个接成更新流，那么，由服务时间流可以生成一个虚拟的更新过程．

最简单的排队系统只有一条服务线，采取先到的顾客先接受服务的原则，并且有足够大的空间(无限制)容纳等待服务的顾客．在时刻t等待服务与正接受服务的顾客数，是一个取非负整值的随机变量，记为 X_t.整值随机过程$\{X_t : t \ge 0\}$称为排队过程，它是连续时间状态离散的随机过程．顾客流与服务时间流都是 Poisson 流(即指数流)的排队过程是最简单的排队过程．由定理 6.１２知道，只有此种情形的排队过程才可能是连续时间的 Markov 链．

排队模型广泛地出现在各个应用领域，如服务系统，交通运输，电脑中信息流的存取，通信系统，商品物流等等．

1．2　排队系统的一般框图,输入过程与输出过程

排队系统的一般框图如下：



值得注意的是，输入过程就是顾客流生成的更新过程，但是输出过程却不是服务时间流生成的虚拟的更新过程．因为输出过程是输入过程与服务时间流共同作用的结果．

记输入过程为 $N_t^{(in)}$（ 即在$(0,t]$中到达的顾客数)，输出过程为 $N_t^{(out)}$（ 即在$(0,t]$中离开系统的顾客数)．那么我们有

$$X_t - X_0 = N_t^{(in)} - N_t^{(out)}. \qquad (7.1)$$

1．3 可逆性引理

如果顾客流是 Poisson 流，且在某个随机的初值X_0时排队过程是可逆的平稳过程，即对于任意$T>0$, 随机过程$\{\hat{X}_t^{(T)} \overset{\Delta}{=} X_{T-t} : t \le T\}$与随机过程$\{X_t : t \le T\}$(前者称为后者的逆过程）有相同的有限维分布族，则输出过程有如下的可逆性引理：

引理 7。1 （可逆性引理） (具 Poisson 输入的可逆排队系统的输出过程)

设排队系统的输入过程是强度为 l 的 Poisson 过程，而且排队过程是可逆的平稳过程，

则　排队系统的输出过程也是强度为 l 的 Poisson 过程.

证明　由于$\{X_t : t \le T\}$每次增加 1 都发生在$t_n \wedge T$时刻，其中$\{t_n\}$为参数为 l 的指数流的累次到达时刻，有可逆性可知$\{\hat{X}_t^{(T)} : t \le T\}$也应具有同样的性质．但是$\{\hat{X}_t^{(T)} : t \le T\}$增加 1 的时刻正是$\{X_t : t \le T\}$减少 1 的时刻，可见$\{X_t : t \le T\}$减少 1 的时刻与$\{t_n\}$同分布．也就是$\{X_t : t \le T\}$减少 1 的时刻也是参数为 l 的 Poisson 流.

[注]　M/M/N 排队系统（参见§2 中的(2．2)段）是满足可逆性引理的典型例子.

2．最简单排队过程—Markov 排队过程

2．1 最简单的排队过程-- $M/M/1$系统

假定服务系统只有一个 "服务员"，顾客到达的时间间隔相互独立同分布，且服从分布 $\exp_l$，而每个顾客接受服务的时间长度也相互独立同分布，服从$\exp_m$分布，并且与排队系统的输入流相互独立．这种系统在排队论中记为 $M/M/1$系统，其中第一个 "M" 表示顾客流服从指数分布，第二个 "M" 表示服务时间长度流服从指数分布，最后一个数字 "1" 表示只有 1 个 "服务员"）

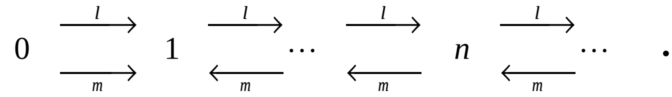
记在$(0,t]$中到达的顾客数目为N_t，则它是强度为 l 的 Poisson 过程．如果在t时刻的排队过程$X_t \ne 0$，则系统不在闲期， 此时接受了服务的顾客的个数(我们将它记为L_t)在很短的时间内，是以强度为m的 Poisson 过程的规律相增加的． 因此，当$X_t \ne 0$且h很小时，在$(t,t+h]$中接受了服务的顾客个数L_h满足

$$P(L_h \ge 2) = o(h)\text{，}\ P(L_h = 1) = mh + o(h)\text{，}\ P(L_h = 0) = (1-mh) + o(h)\text{．} \quad (7.2)$$

(这里需要警觉的是： 接受了服务的顾客个数L_t，并不总是顾客接受服务的时间长度流对应的 Poisson 过程$N_t^{(serve)}$，这是因为当队伍空时就不会出现服务．虽然服务流的计数过程$\{N_t^{(serve)} : t \ge 0\}$总是与顾客流的计数过程$\{N_t^{(in)} : t \ge 0\}$独立的，但是$\{L_t : t \ge 0\}$因为受排队过程$\{X_t : t \ge 0)$是否处于忙期的影响，从而也受$\{N_t^{(in)} : t \ge 0\}$的影响．所以，要想用服务流的计数过程$N_h^{(serve)}$来描述接受了服务的顾客个数，必须满足： $X_t \ge$服务线的数目，且h很小．

此系统的排队过程X_t的状态空间为$S = \{0,1,\cdots,n,\cdots\}$．下面考察排队过程在$(t,t+h]$

中的转移情况．与例６．３６类似，对转移速率$q_{ij},(i \neq j)$我们有如下的流向图：



由此得到排队过程X_t是转移速率矩阵为

$$\mathbf{Q}=\begin{pmatrix} -l & l & & & \\ m & -(l+m) & l & & \\ & m & -(l+m) & l & \\ & & \ddots & \ddots & \ddots \\ & & & & \end{pmatrix}. \qquad (7.3)$$

X_t是连续时间的 Markov 链． 即是具有常数生长率l与常数死亡率的单侧生灭过程．当$m > l$时，它有可逆分布，此可逆分布是参数为$\frac{l}{m}$的几何分布(参见下一段)

$$k = (p_0,\cdots,p_n,\cdots)\text{，} p_n = (1-\frac{l}{m})(\frac{l}{m})^n. \qquad (7.4)$$

M/M/1 可逆系统在稳定时的平均排队长度，停留时间分布与服务员的等待时间的分布

当$m > l$时，$M/M/1$ 系统为可逆的，在系统达到稳定时有

(1) 平均排队长度为

$$L=\sum_k kp_k=\sum_k k(1-\frac{l}{m})(\frac{l}{m})^k=\frac{l}{m}(1-\frac{l}{m})[\frac{d}{dx}(\frac{1}{1-x})]_{x=\frac{l}{m}}=\frac{l}{m-l}. \qquad (7.5)$$

(2) 停留时间T的分布为$\exp_{m-l}$

为了验证这个事实，我们用随机变量V来记排队系统达到稳定时的队伍长度．那么

$$V \sim \begin{pmatrix} 0 & \cdots & n & \cdots \\ 1-\frac{l}{m} & \cdots & (\frac{l}{m})^n(1-\frac{l}{m}) & \cdots \end{pmatrix}.$$

一个顾客进入系统后的停留时间T是随机变量．当排队系统达到稳定时，$P(T \le t \mid V=k)$是$k+1$个顾客 (系统中的k个顾客与进入系统的顾客)连续接受服务的时间的分布函数，也就是$k+1$个参数为m的独立的指数分布随机变量的和的分布函数，故而应该是$\Gamma(k+1,l)$的分布函数． 于是

$$P(T \le t)=\sum_{k=0}^{\infty} P(T < t \mid V=k)P(V=k)$$

$$=\sum_{k=0}^{\infty}\int_0^t \frac{m^{k+1}}{k!}s^k e^{-ms}ds(1-\frac{l}{m})(\frac{l}{m})^n=\int_0^t (m-l)e^{-(m-l)s}ds. \qquad (7.6)$$

即 $T\sim \exp_{m-l}$. **这个结论非常符合直观的印象.**

(3) **服务员的等待时间间隔**S **的分布为** $\exp_l$ **.**

事实上，服务员的等待时间的间隔 S，**是一个顾客离开系统时与前面的一个顾客离开系统的时间的差，它也是随机变量．设前面的一个顾客离开系统时系统中的顾客数为**V'．**那么在排队系统达到稳定时**V'**与**V **同分布．这时** $P(S\le t\,|\,V'>0)$ **应该是** $\exp_m$ **的分布函数，而** $P(S\ge t\,|\,V'=0)$ **应是独立的** $\exp_l$ **随机变量与** $\exp_m$ **随机变量的和的分布函数．即**

$$P(S\le t\,|\,V'=0)=\int_0^t\int_0^s l\cdot e^{-lu}m\cdot e^{-m(s-u)}duds=\int_0^t l\cdot e^{-ms}\frac{m}{m-l}(e^{(m-l)s}-1)ds$$

$$=\frac{m}{m-l}(1-e^{-l\cdot t})-\frac{l}{m-l}(1-e^{-mt})=1-\frac{1}{m-l}(m\cdot e^{-l\cdot t}-l\cdot e^{-m\cdot t}).$$

于是由全概率公式得到

$$P(S\le t)=P(S\le t\,|\,V'>0)P(V'>0)+P(S\le t\,|\,V'=0)P(V'=0)$$

$$=(1-e^{-mt})\frac{l}{m}+[1-\frac{1}{m-l}(m\cdot e^{-l\cdot t}-l\cdot e^{-m\cdot t})](1-\frac{l}{m})=1-e^{-l\cdot t}. \qquad (7.7)$$

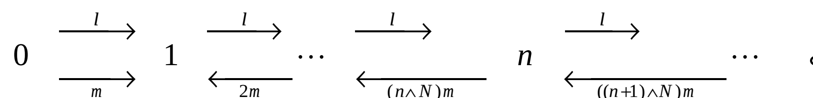
可见 $S\sim\exp_l$.

2．2 N **个服务员的简单排队过程 —** $M/M/N$ **系统**

1．转移速率矩阵

假定服务系统有 N **个 ”服务员”，他们分别独立地服务．顾客到达的间隔仍为参数** l **的独立指数分布，且每个顾客接受服务的时间长度仍服从与之独立的，参数为** m **的独立指数分布．这种系统在排队理论中记为** $M/M/N$ **系统，前两个字母表示指数分布，最后一个数字“**N**”表示有** N **个“服务员”．**

排队过程 X_t **的状态空间仍是** $S=\{0,1,\cdots,n,\cdots\}$**．　可以用如下的流向图来分析转移速率** $q_{ij},(i\ne j)$**：**



由此流向图可以得到

$$\mathbf{Q}=\begin{pmatrix} -l & l & & & & \\ m & -(l+m) & l & & & \\ & \ddots & \ddots & \ddots & & \\ & & (n\wedge N)m & -[(n\wedge N)+l] & l & \\ & & & \ddots & \ddots & \end{pmatrix}. \qquad (7.8)$$

它是互通的，且具有配称列 $m=(m_0,m_1,\cdots)$，其中

$$m_0=1,\ m_n=\frac{l}{(n\wedge N)m}m_{n-1}=\frac{1}{\prod_{k=1}^{n}(k\wedge N)}\left(\frac{l}{m}\right)^{n+1}. \qquad (7.9)$$

于是存在可逆分布当且仅当 $l<m$．在条件成立时，令

$$p=\frac{1}{\sum_{k=0}^{N}\frac{1}{k!}(\frac{l}{m})^k+\frac{1}{N!}\sum_{j=0}^{\infty}\frac{1}{N^j}(\frac{l}{m})^{N+j}}\ m. \qquad (7.10)$$

那么 p 就是可逆分布． 因此排队过程 X_t 是连续时间的 Markov 链，在 $l<m$ 时其转移矩阵

$\boldsymbol{P}(t)$ 有极限，且

$$\boldsymbol{P}(t)\xrightarrow{t\to\infty}\mathbf{1}^T p\ .$$

在达到稳定时，排队系统处于闲期的概率为 p_0，处于忙期的概率为 $1-p_0$．

由此我们也可以得到，在 $m>l$ 时 $M/M/N$ 系统达到稳定时的平均排队长度．

2．可逆 $M/M/N$ 系统的特性

命题 7．2　若 $M/M/N$ 排队过程的初值为可逆分布(或者任意初值，并经过时间充分长后)，那么

(1) 正在系统中的人数与输出过程的过去情况相互独立．

(2) 给定顾客在此系统中所停留的时间，与他离开前的输出流独立．

证明　(1) 输出过程的过去就是逆过程的输入过程的将来，这正是指数流的将来．由指数流的性质，它应与在系统中的人数独立．

(2) 由引理７．１，逆过程的输入过程是指数流．再用（１），逆过程在进入系统时刻以后的输入流就与已进入的顾客在系统中的停留时间独立．这就是说，顾客离开系统前的输出流与他在系统中的停留时间独立．

3． 可逆 $M/M/N$ 系统在稳定时的平均等待时间

设 $m>l$．在达到稳定后(时间充分长,致使系统近似平稳)，对于在时刻 t 来到排队系统的顾客的平均等待时间，可以分析如下． 由于排队系统平稳，对于在系统中的顾客(包括在排队的与正在接受服务的顾客)数 X_t，应有 $P(X_t=i)=p_i$．记时刻 t 来到的顾客等待时间

为W．于是当$i<N$时有

$$P(W=0\mid X_t=i)=1.$$

而当$i\geq N$时，此顾客的等待时间恰为t后，第$i-N+1$个顾客离开的时间（因为在这之前，N条服务线忙于为排队在为前面的$i-N+1$个顾客服务，只有当第$i-N+1$个顾客结束服务后，这N条服务线中才有一条空出来）．记t时间以后第k个顾客离开的时间为$t+s_k$，注意顾客与服务线的服务时间是相互独立的，当$i\geq N$时就有

$$P(s_1\geq s\mid X_t=i)=P(N\text{条服务线无一在}s\text{前结束}\mid X_t=i)=(e^{-m\cdot s})^N=e^{-Nm\cdot s}.$$

可见s_1服从参数为Nm的指数分布．再则，在$i\geq N+1$时，当第1个顾客结束服务后，这N条服务线各自需要多少时间结束已开始的服务，是与此前的情况相互独立的，而且仍相互独立地服从参数为m的指数分布。所以，当$i\geq N+k-1$时，$\{s_j-s_{j-1}\}_{1\leq j\leq k}$服从相互独立的指数分布$\exp_m$．故而 $s_k\sim\Gamma(k,Nm)$分布，即

$$P(s_k\geq s)=\int_s^\infty Nm\frac{(Nm\cdot x)^{k-1}}{(k-1)!}e^{-Nm\cdot x}dx=\sum_{j=0}^{k-1}e^{-Nm\cdot s}\frac{(Nm\cdot s)^j}{j!}.$$

最后，原来排队的i人在时刻t新来的顾客加入后成为$i+1$人，除去在接受服务的N人外，需要等待$i+1-N$次服务，因此由p_i的表达式我们得到：

$$P(W\geq s)=\sum_{i=N}^\infty p_i P(s_{i+1-N}\geq s)=\sum_{i=N}^\infty p_N(\frac{l}{Nm})^{i-N}\sum_{j=0}^{i-N}e^{-N\cdot m\cdot s}\frac{(N\cdot m\cdot s)^j}{j!}$$

$$=p_N e^{-N\cdot ms}\sum_{l=0}^\infty(\frac{l}{Nm})^l\sum_{j=0}^{l}\frac{(Nms)^j}{j!}=p_N e^{-Nms}\sum_{j=0}^\infty\frac{(Nms)^j}{j!}\sum_{l=j}^\infty(\frac{l}{Nm})^l$$

$$=p_N e^{-Nm\cdot s}\sum_{j=0}^\infty(\frac{l}{Nm})^j\frac{(Nms)^j}{j!}\frac{1}{1-\frac{l}{Nm}}=p_N e^{-(Nm-l)s}\frac{Nm}{Nm-l}.$$

于是有下面的定理

定理７.３　可逆$M/M/N$系统在稳定时有

1#　需要等待的概率为：

$$P(W>0)=p_N\frac{Nm}{Nm-l}. \tag{7.11}$$

2#　平均等待时间为

$$EW=\int_0^\infty P(W>s)ds=p_N\frac{Nm}{(Nm-l)^2}. \tag{7.12}$$

特别当$N=1$时有 $p_1=\frac{l}{m}(1-\frac{l}{m})$，$EW=\frac{l}{m}\frac{1}{m-l}$. 同时也可得平均停留时间$T$，因为它是等待时间$W$与服务时间的和.

*2．3 序贯排队与排队网络系统

1．简单的序贯排队模型

设排队系统由两个$M/M/1$子系统串联而成．设进入第一个子系统的顾客流为参数 l 的指数流，服务时间流为与之独立的,参数为m_1的指数流．从第一个子系统离开的顾客进入第二个子系统，其服务时间流是与前两个流独立的参数为m_2的指数流．假定$m_1 \wedge m_2 > l$．

这时，第一个子系统有可逆分布，它的排队过程是空间齐次的单侧生灭过程，而且其转移矩阵的每一行都趋于其可逆分布．于是不管什么初值，只要时间充分长，就可以认为可逆性引理７．１的条件成立． 由引理 7.1，第一个子系统的输出过程(也就是第二个子系统的输入过程)也是以 l 为强度的 Poisson 过程， 并且第二个子系统也有同样的性质．

2．简单的排队网络

假定服务网络系统设有n个服务点，每个服务点设置一条服务线，分别独立地服务．假定在第i $(i \le n)$个服务点所需的服务时间为参数m_i的指数流。 令$\mathbf{P}=(p_{ij})$为一个随机矩阵，其中p_{ij}在$i \ne j$时表示：第i个服务点接受完服务后的顾客，转而再去第j个服务点的概率．而p_{ii}表示顾客在第i个节点（服务点）接受完服务后，离开此服务网络系统的概率．时刻t在第i个服务点的排队过程为$X_t^{(i)}$．定义$X_t \overset{\Delta}{=} (X_t^{(1)},\cdots,X_t^{(n)})$，这个$n$维过程称为在此服务网络系统中的排队过程．假定由系统外进入系统各个服务点的顾客流，是彼此相互独立的，而且进入第i个服务点的顾客流为参数 $l_i^{(0)}$ 的指数流。 再假定m_i比较大，以使每个服务点上的排队过程是可逆的（在后面我们将会看到它们成立的条件）．

由于顾客在各个服务点间有转移，所以进入第i个服务点的实际强度（记为 l_i）要比 $l_i^{(0)}$ 大．由引理 7.1，只要在第i个服务点的排队过程是可逆的，那么在时间充分长以后，在这个服务点的输出流也应该是参数为 l_i 的指数流．

记第 j 个节点(服务点)接受到的外来顾客的时间间隔流为$\{T_k^{(0\to j)}\}$，而从节点i接受到的顾客的时间间隔流为$\{T_k^{(i\to j)}\}$。 由 Poisson 过程的分流定理知道，$T_k^{(i\to j)} \sim \exp_{p_{ij}l_i}$．而且对于固定的$k$，$T_k^{(0\to j)},T_{k-1}^{(1\to j)},\cdots,T_{k-1}^{(n\to j)}$相互独立．这使得在节点$j$实际接受到的顾客的时间间隔为

$$T_k^{(0\to j)} \wedge T_{k-1}^{(1\to j)} \wedge \cdots \wedge T_{k-1}^{(n\to j)} \sim \exp_{l_j^{(0)}+\sum_i p_{ij}l_i}. \qquad (7.13)$$

于是顾客流的强度 l_j 应该满足流方程

$$l_j = l_j^{(0)} + \sum_i p_{ij} l_i . \tag{7.14}$$

这个网络排队过程转移速率矩阵 $\mathbf{Q}$ 较为复杂，但是 $\mathbf{Q}$ 的流向却可以简单表示为:

$$\cdots(k_1,\cdots,k_i-1,\cdots,k_n) \underset{m_i}{\overset{l_i}{\rightleftarrows}} (k_1,\cdots,k_i,\cdots k_j,\cdots,k_n) \underset{m_j}{\overset{l_j}{\rightleftarrows}} (k_1,\cdots,k_i,\cdots,k_j+1,\cdots,k_n)\cdots,$$

即

$$q_{(k_1,\cdots,k_i\cdots,k_n),(k_1,\cdots,k_i+1,\cdots,k_n)} = l_i,\ \ q_{(k_1,\cdots,k_i+1\cdots,k_n),(k_1,\cdots,k_i,\cdots,k_n)} = m_i . \tag{7.15}$$

现在我们假定流方程的解 $\{l_1,\cdots,l_n\}$ 满足：$l_i < m_i,(i=1,\cdots,n)$。 那么仿照第 6 章中的讨论，就可以得到 $\mathbf{Q}$ 有配称列 $\text{p} = (\cdots,p_{(k_1,\cdots,k_i\cdots,k_n)},\cdots)$，它是概率向量，其分量为

$$p_{(k_1,\cdots,k_i\cdots,k_n)} = \prod_{i=1}^{n} \left(\frac{l_i}{m_i}\right)^{k_i} \left(1-\frac{l_i}{m_i}\right). \tag{7.16}$$

最后， 用定理 6.26 就得到以 $\mathbf{Q}$ 为转移速率的，时间连续的 Markov 链的转移矩阵 $\mathbf{P}(t)$ 满足

$$\mathbf{P}\ \ (t) \to \mathbf{1}^T \text{p}, \quad (t\to\infty),$$

并且遍历定理成立. 由(7.16)可以看出，在各个服务点的排队过程是渐近地相互独立的.

2．4 M／M／∞排队系统

这时有 ∞ 条彼此独立地工作的服务线．排队过程 X_t 的转移速率矩阵为

$$\mathbf{Q} = \begin{pmatrix} -l & l & & & & \\ m & -(l+m) & l & & & \\ & \ddots & \ddots & \ddots & & \\ & & Nm & -(l+Nm) & l & \\ & & & \ddots & \ddots & \ddots \\ & & & & & \end{pmatrix}, \tag{7.17}$$

它是互通的，具有可逆分布

$$p_N = e^{-\frac{l}{m}} \left(\frac{l}{m}\right)^N . \tag{7.18}$$

而且也有

$$\mathsf{P}(t) \xrightarrow{t\to\infty} \mathbf{1}^T p \quad .$$

下面我们进一步求 X_t 的分布 $p_k(t) = P(X_t = k)$ 的表达式.

这时的 Master 方程

$$(p_0'(t), p_1'(t), \cdots) = (p_0(t), p_1(t), \cdots)\ \mathbf{Q}$$

的分量形式为

$$\begin{cases} p_0'(t) = -l p_0(t) + m p_i(t) \\ p_k'(t) = l p_{k-1}(t) - (l + km) p_k(t) + (k+1) m p_{k+1}(t) \end{cases} . \qquad \text{(7．19)}$$

这组方程等价于 X_t 的矩母函数

$$G(t,z) \stackrel{\Delta}{=} E z^{X_t} = \sum_{k=0}^{\infty} p_k(t) z^k \qquad \text{(7．20)}$$

满足如下的偏微分方程：

$$\frac{\partial G}{\partial t} = \sum_{k=0}^{\infty} p_k'(t) z^k = (z-1)(lG - m\frac{\partial G}{\partial z}) .$$

即

$$\frac{\partial G}{\partial t} + m(z-1)\frac{\partial G}{\partial z}) - l(z-1)G = 0 . \qquad \text{(7．21)}$$

作变换 $G(t,z) = e^{\frac{l}{m}(z-1)(1-e^{-mt})} H(t,z)$**，就简化为**

$$\frac{\partial H}{\partial t} + m(z-1)\frac{\partial H}{\partial z} = 0 . \qquad \text{(7．22)}$$

两边乘以 e^{-mt} **，记函数对** $(H(t,z)$ **，** $(z-1)e^{-mt})$ **关于** (t,z) **的 Jacobian 行列式为**

$\dfrac{\partial(H,(z-1)e^{-mt})}{\partial(t,z)}$**，那么方程（7．22）就可改写成**

$$\frac{\partial(H,(z-1)e^{-mt})}{\partial(t,z)} = 0 . \qquad \text{(7．22)'}$$

这就说明了 $H(t,z)$ **与** $(z-1)e^{-mt}$ **函数相关．即存在一个函数** $h(x)$ **使**

$$H(t,z) = h((z-1)e^{-mt}) . \qquad \text{(7．23)}$$

设t＝0时系统中的顾客数 X_0 **的分布为** $P(X_0 = k) = a_k$**．那么**

$$\sum_{k=0}^{\infty} a_k z^k = G(0,z) = H(0,z) = h(z-1) .$$

于是

$$h(z) = \sum_{k=0}^{\infty} a_k (z+1)^k .$$

因此

$$G(t,z)=e^{\frac{l}{m}(z-1)(1-e^{-mt})}\sum_{k=0}^{\infty}a_k(1+(z-1)e^{-mt})^k\ .\qquad (7.24)$$

显见 $G(t,z)$ 的表达式完全依赖于初始资料 $G(0,z)$ 的选取．当取 $G(0,z)=z^i$ 时，我们把得到的 $G(t,z)$ 记为

$$G_i(t,z)\overset{\Delta}{=}\sum_{j=0}^{\infty}p_{ij}(t)z^j\ .$$

由此可得下面的定理

定理７.４　可逆 $M/M/\infty$ 系统排队过程的的转移函数为

$$p_{ij}(t)=e^{-\frac{l}{m}(1-e^{-mt})}\sum_{k=0}^{i\wedge j}C_k^i\left(\frac{l}{m}\right)^{j-k}\frac{e^{-kmt}(1-e^{-mt})^{i+j-2k}}{(j-k)!}.\qquad (7.25)$$

于是

$$p_{0j}(t)=e^{-\frac{l}{m}(1-e^{-mt})}\frac{[\frac{l}{m}(1-e^{-mt})]^j}{j!}.\qquad (7.26)$$

即 当 $X_0=1$ 时，$X_t\sim\exp_{\frac{l}{m}(1-e^{-mt})}$．由此可见当 $t\to\infty$ 时 X_t 有极限分布 $\exp_{\frac{l}{m}}$，这就再一次求得了 X_t 的不变分布．

［注 1］以上的方法与结论可以推广到λ依赖于 t 的情形,只要假定 λ(t)≤常数λ$_0$．这时有

$$\lim_{t\to\infty}\left(P(X_t=k)-e^{-\Lambda(t)}\frac{\Delta(t)^k}{k!}\right)=0\ ,$$

其中

$$\Lambda(t)=\frac{1}{m}\int_0^t l(s)ds\ .$$

即当 t 非常大的时候，排队过程 X_t 的分布与 $\exp_{\Delta(t)}$ 非常接近．

［注 2］（成批顾客的排队系统）

上面的方法和结论，还可以推广到顾客有成批到达的情形．假定顾客在参数为λ的 Poisson **过程的跳跃时刻上到达，但到达的顾客是成批的，即到达人数 u 是一个非负整值的随机变量，其母函数记为**

$$B(z)=\sum_{k=0}^{\infty}b_kz^k,(b_k\overset{\text{D}}{=}P(u=k)).\qquad (7.27)$$

又假定服务时间仍为参数为 m 的独立同分布的指数分布．那么用与推导 $M/M/\infty$ 系统的排队过程的矩

母函数相似的方法，对于这里的成批排队系统的排队过程 X_t 的矩母函数

$$G(t,z)\overset{\Delta}{=}Ez^{X_t}=\sum_{k=0}^{\infty}p_k(t)z^k,$$

可以得到

$$G(t,z)=G(0,1-(1-z)e^{-mt})e^{\frac{1}{m}\int_{1-(1-z)e^{-mt}}^{z}\frac{1-B(s)}{1-s}l\,(t+\frac{1}{m}\ln\frac{1-s}{1-t})ds}. \qquad (7.28)$$

3. 排队系统的一般概念

3.1 关于排队论的一般注记

排队系统是服务系统,交通运输,通讯系统等许多领域中的问题简化后的数学模型.

由第 2 节可以看出，即使是很简单的排队系统的数学模型，其数学处理的也是很复杂的.应用问题中的需要与数学的复杂性的矛盾在排队系统中突出地尖锐.于是，不同的实际工作者在各自的范围使用自己的直观近似.对于同样问题的看法也不尽一致.至今关于排队论的文献已超过 10 万，而且每年还以近两千篇的数量增加．本章只是介绍一点简单的框架.

排队理论涉及了一个复杂系统，它包含：

1．输入过程：它一般是一个更新流，称为"顾客"流．更广地，也可以是每次批量输入的输入流.

2．服务设备(服务线或服务员)的数目：可以有 1 个、*N*个或∞个．它们之间是彼此独立地工作的.

3．服务设备对于不同顾客的服务时间是独立同分布的随机变量，它们与输入过程独立.

4．服务规则：最常见的是先来先服务(记为 FIFO，即 First In First Out).
除单个服务外还有成批服务(顾客够一定数才启动服务).顾客接受服务又可分三种方式：

第一种是等待制：在服务设备全用上时，多余的顾客排队等候,这需要有充分的等候设施，这称为有无限空间等待的情形；

第二种是消失制：顾客见到所有服务设备都在忙就离去,例如电话通讯；

第三种是有限制的排队，即每个顾客看到队伍长度超过了某个预定的 N 则离去(等候设备有限制).

5．排队系统研究的主要指标有：

(1) 稳态时排队长度的不变分布与其平均值，方差(在等待制时用作设计等候室的参考).

(2) 设备在平稳状态下的平均忙期长度、平均闲期长度.

(3) 稳定时拒绝服务的概率(消失制).

(4) 稳定时等待时间的分布与平均值,方差 (等候制时用以设置服务线).

(5) 稳定时的效率(最高服务率时平均能服务的顾客数).

6．一个排队系统用下面的记号表示：输入流分布／服务流分布／服务线数目.

M：表示指数分布.

E_k：表示 k 阶 Erlang 分布，即 $\Gamma(k,l)$ 分布.

G：表示一般分布函数，记为 $G(x)$，在输入流中 G 有时也用 GI 表示.

D：表示流为常值.

常见的有排队系统有：$M/M/1$，$M/M/N$ ，$M/M/\infty$，$G/M/1$ ，$M/G/1$，$GI/G/1$ (也记为$G/G/1$)等等.

7．描述排队系统的一个最基本的量是排队过程 X_t，即时刻 t 在系统中的顾客数(包括正在被服务的顾客与排队的顾客)， 这是一个连续时间的取非负值的随机过程. 除 $M/M/*$ 系统以外，一般的排队过程 X_t 并不是 Markov 链. 但是，有时加入一些由附加信息提供的辅助随机过程后,可以使它们合起来的向量过程有 Markov 性. 也就是使排队过程成为某个高维 Markov 链的一个分量.

8．由于在文献中已经证明了：在 $[0,\infty)$ 上具有密度的随机变量的任意的分布函数，都可以用 Erlang 分布的分布函数的混合来近似. 因此在实用中,总认为可以用混合 Erlang 流近似任意更新流.

3．2 $M/M/N$ 消失制

消失制情形时没有得到服务的顾客都自动离去. 所以此时的排队人数就是正在接受服务的人数. 因此排队的队伍长度不超过 N . 此时排队过程 X_t 的转移速率矩阵为

$$\mathbf{Q}=\begin{pmatrix} -l & l & & & & \\ m & -(l+m) & l & & & \\ & 2m & -(l+2m) & l & & \\ & & \ddots & \ddots & \ddots & \\ & & & (N-1)m & -(l+(N-1)m) & l \\ & & & & Nm & -Nm \end{pmatrix}.$$

它是互通的，而且有可逆分布(这个稳态分布的公式，称为 Erlang 公式)

$$p=\left(\sum_{j=0}^{N}\frac{1}{j!}\left(\frac{l}{m}\right)\right)^{-1}\left(1,\frac{l}{m},\cdots,\left(\frac{l}{m}\right)^{N}\right). \qquad (7．29)$$

此时仍有

$$\mathsf{P}(t)\xrightarrow{t\to\infty}\mathbf{1}^{T}p$$

这时，在稳态时新来的顾客被拒绝，当且仅当在系统中的顾客数为 N . 因而我们得到下面的定理

定理7．5 对于可逆 $M/M/N$ 系统，在稳定时新来的顾客被拒绝的概率为

$$p_N=\left(\sum_{j=0}^{N}\frac{1}{j!}\left(\frac{l}{m}\right)\right)^{-1}\left(\frac{l}{m}\right)^{N}. \qquad (7．30)$$

3．3 $M/G/1$ 排队系统

如果服务时间流不是指数流，而是一般的更新流。即服务时间为独立同分布，其分布函数为 $G(t)$，或分布密度为 $g(x)$ （最简单的情形是 $G(t)$ 为 $Erlang(l,k)$ 分布），其它假定都与 $M/M/1$ 相同。 这样的排队系统就是 $M/G/1$.

系统 $M/G/1$ 的排队过程 X_t，不再是时间连续的 Markov 链. 这是因为服务时间不是

指数流，即更新间隔的分布不是无记忆的．所以，正在接受服务的顾客已经在服务线上已花费的时间，就应该是一个需要记忆的参数．在次参数给定时，系统中的排队人数也不具有 Markov 性．但是，如果只考虑顾客离开的时刻 s （注意它是随机的）， 由于此时正好完成了一个服务周期，于是在这些时刻排队人数的变化只由输入决定，而输入是指数流，我们可以猜想排队过程在这些时刻上具有 Markov 性.即只考虑排队过程在顾客离开的时间列 $\{s_n\}$ 上就是离散时间的 Markov 链．我们把它表达为下面的定理

定理７.６ $M/G/1$ 系统的排队过程 X_n 限制在顾客离开的时间列 s_n 上，得到的 X_{s_n} 是时间离散的 Markov 链，它恰好记录了排队过程各个顾客刚离开后的排队队伍长度，其转移矩阵为

$$\mathbf{P}=\begin{pmatrix} a_0 & a_1 & a_2 & a_3 & \cdots \\ a_0 & a_1 & a_2 & a_3 & \cdots \\ 0 & a_0 & a_1 & a_2 & \cdots \\ 0 & 0 & a_0 & a_1 & \cdots \\ \vdots & \vdots & \ddots & \ddots & \ddots \end{pmatrix}, \qquad （7．31）$$

其中

$$a_j == \int_0^\infty \frac{(l\cdot t)^j}{j!} e^{-lt} g(t)dt\,. \qquad （7．32）$$

这个 Markov 链没有可逆分布．

证明　为了使讨论简单，我们不妨假定 $G(t)$ 具有分布密度 $g(t)$．从直观上看，把第 n 个顾客离开的时刻记为 σ_n， 而记此时在系统中还留下的排队人数为 Z_n． Z_n 是由此顾客在接受服务期间进入系统的（且与此顾客的服务独立的）指数顾客流决定的．我们把第 $n-1$ 个顾客离开（即第 n 个顾客接受服务）至第 n 个顾客离开期间(即 $(\sigma_{n-1},\sigma_n]$，它恰为一个服务周期)进入系统的顾客数记为 Y_{n-1}．那么，Y_{n-1} 与 $(Z_{n-1},\cdots,Z_1)$ 独立．同时我们有

$$Z_n=\begin{cases} Y_{n-1}+Z_{n-i}-1 & (\text{若}Z_{\text{n-1}}>0) \\ Y_{n-1} & (\text{若}Z_{n-1}=0) \end{cases}.$$

在 $i>0$ 时推得

$$P(Z_n=j\,|\,Z_{n-1}=i,Z_{n-2}=i_{n-2},\cdots)=P(Y_{n-1}=j-i+1\,|\,Z_{n-1}=i,Z_{n-2}=i_{n-2},\cdots)$$

$$=P(Y_{n-1}=j-i+1)=P(Z_n=j\,|\,Z_{n-1}=i)\,.$$

而在 $Z_n=0$ 时，Markov 性是显然的． 这就证明了 $\{Z_n\}$ 是 Markov 链．我们进一步求它的转移矩阵．（1）当 $j<i-1$ 时，显见 $P(Z_n=j\,|\,Z_{n-1}=i)=0$．

（2）在 $j\ge i-1, i>0$ 的情形．由于 $(\sigma_{n-1},\sigma_n]$ 对应一个服务周期，而

$Y_{n-1}=N_{s_n}-N_{s_{n-1}}$，$s_n-s_{n-1}\sim g(t)$，我们用全期望公式求得转移概率

$$P(Z_n=j\,|\,Z_{n-1}=i)=P(Y_{n-1}=j-i+1)$$

$$=\int_0^\infty P(N_{s_n}-N_{s_{n-1}}=j-i+1\,|\,s_n-s_{n-1}=t)g(t)dt\overset{*}{=}\int_0^\infty P(N_t=j-i+1\,|\,T'=t)g(t)dt$$

$$=\int_0^\infty P(N_t=j-i+1)g(t)dt=\int_0^\infty \frac{(l\cdot t)^{j-i+1}}{(j-i+1)!}e^{-lt}g(t)dt\text{.}$$

(注 *的严格推导要用 Poisson 过程的强再生性).

对$i=0$的情形，类似地可算出

$$P(Z_n=j\,|\,Z_{n-1}=0)=\int_0^\infty \frac{(l\cdot t)^{j}}{j!}e^{-lt}g(t)dt=a_j\text{.} \qquad 】$$

下面我们来求它存在不变分布的条件.

记平均服务时间为v （它相当于$M/M/1$ 系统中的$\frac{1}{m}$)，那么

$$v=\int_0^\infty tg(t)dt\text{.}$$

注意，转移矩阵 P 的首行是一个分布，记此分布的期望为r，则

$$r=\sum_{j=0}^\infty ja_j=\int_0^\infty j\sum_{j=1}^\infty\frac{(l\cdot t)^j}{j!}e^{-l\cdot t}g(t)dt=l\int_0^\infty g(t)dt=l\cdot v\text{.}$$

受$M/M/1$ 系统的启发，我们猜测并证明：当$l<\frac{1}{n}$时，即$r<1$ 时，$\{Z_n\}$存在唯一不变分布$p=(p_0,p_1,\cdots)$. 为此我们只须求出其母函数 $p(z)=\sum_{j=0}^\infty p_jz^j$. 记

$A(z)=\sum_{j=0}^\infty a_jz^j$. 由 P 的形式,用$p=p$ P 的分量形式

$$p_j=p_0a_j+\sum_{i=1}^{j+1}p_ia_{j-i+1}\text{,}$$

便得

$$p(z)=p_0A(z)+\sum_{j=0}^\infty\sum_{i=1}^{j+1}p_ia_{j-i+1}z^j\text{,}$$

其中后一个和数为

$$\frac{1}{z}\sum_{i=1}^{\infty}p_i z^i \sum_{j=i-1}^{\infty}a_{j-i+1}z^{j-i+1} = \frac{(p(z)-p_0)A(z)}{z}.$$

最后得到

$$p(z) = \frac{(z-1)p_0 A(z)}{z - A(z)}.$$

由此得 $p(1) = \frac{p_0}{1-r}$．故而 $p_0 = 1 - r$．于是当 $r < 1$时，不变分布唯一地存在．容易验证它是互通的．由定理５．５０知道$\{Z_n\}$是正常返的，易见它还是非周期的．所以由该定理后面的注 1，有

$$\mathbf{P}^n \to \mathbf{1}^T p \ .$$

从而p 代表了系统稳定时，在顾客离开系统时，还留在系统中的顾客数的分布． 综上得到下面的定理

定理７．７ 对于排队系统$M/G/1$，在满足条件$\int_0^{\infty} tg(t)dt < \frac{1}{l}$时，排队过程在顾客离开时刻上导出的离散时间 Markov 链存在唯一的不变分布p ，它的矩母函数为

$$p(z) = \frac{(z-1)(1-r)A(z)}{z-A(z)}, \qquad （７．３３）$$

其中

$$A(z) = \sum_{j=0}^{\infty} a_j z^j, \quad r = l\int_0^{\infty} tg(t)dt. \qquad （７．３４）$$

并且满足 $\mathbf{P}^n \to \mathbf{1}^T p$.】

经过较多的计算，还可以得到$M/G/1$系统的平均忙期为

$$\int_0^{\infty} \sum_{n=1}^{\infty} e^{-lt}\frac{(l\cdot t)^{n-1}}{n!} g^{*n}(t)dt, \qquad （７．３５）$$

其中g^{*n}为g 的n次卷积.

3．4 $G/M/1$排队系统

如果服务时间服从参数为m 的独立指数分布，而顾客到达的流是一个一般的更新流(即间隔时间为独立同分布的分布函数$G(t)$，或分布密度$g(t)$)，其它假定都与$M/M/1$相同，则这样的排队系统记为$G/M/1$．这个系统中顾客到达时间的间隔并不服从指数分布，因此时间间隔的分布不是无记忆的，也就是说，系统等待顾客到达所花费的时间，是一个需要记忆的参数．在此参数给定时，系统中的排队人数不再具有 Markov 性．但是，如果限制于考虑顾客到达的时刻t_n，由于此时正好完成了一个完整的“到达周期”， 我们可以

猜测 排队过程在时间列t_n上有 Markov 性，这就是下面的定理

定理７．８ $G/M/1$系统的排队过程X_t限制在顾客的到达的时刻t_n上，得到的X_{t_n}是时间离散的 Markov 链，它恰好记录了排队过程各个顾客刚到时所看到的排队队伍长度，其转移矩阵为

$$=\begin{pmatrix} \sum_{k=1}^{\infty} b_k & b_0 & 0 & 0 & \cdots \\ \sum_{k=2}^{\infty} b_k & b_1 & b_0 & 0 & \cdots \\ \sum_{k=3}^{\infty} b_k & b_2 & b_1 & b_0 & \cdots \\ \sum_{k=4}^{\infty} b_k & b_3 & b_2 & b_1 & \ddots \\ \vdots & \vdots & \ddots & \ddots & \ddots \end{pmatrix}, \quad (7.36)$$

其中

$$b_j == \int_0^{\infty} \frac{(m\cdot t)^j}{j!} e^{-m.t} g(t)dt. \quad (7.37)$$

证明 与 M/G/1 排队系统对偶地，令

$\hat{Z}_n$=第n个顾客来到系统时见到在系统中的顾客数.

把区间$(t_{n-1}, t_n]$中离开排队系统的顾客个数记为$\hat{Y}_{n-1}$．那么

$$\hat{Z}_n = \max(\hat{Z}_{n-1}+1-\hat{Y}_{n-1}, 0).$$

与 M/G/1 系统类似地可以证明，$\hat{Z}_n$是 Markov 链．服务时间指数流对应的更新过程为 Poisson 过程记为N'_t．又设$\{T_n\}$是一个以$g(t)$为密度的更新流，且$t_n = \sum_{k=1}^{n} T_k$．我们用全期望公式求得$\{\hat{Z}_n\}$的转移概率如下：

(1) 在$i+1 \geq j > 0$时，区间$(t_{n-1}, t_n]$中并未出现闲期，所以

$$P(\hat{Z}_n = j \mid \hat{Z}_{n-1} = i) = P(\hat{Y}_{n-1} = i-j+1)$$

$$= \int_0^{\infty} P(N'_{t_n} - N'_{t_{n-1}} = i-j+1 \mid T_n = t) g(t)dt = \int_0^{\infty} P(N_t' = i-j+1 \mid T_n = t) g(t)dt$$

$$= \int_0^{\infty} P(N_t' = i-j+1) g(t)dt = \int_0^{\infty} \frac{(m\cdot t)^{i-j+1}}{(i-j+1)!} e^{-lt} g(t)dt, (0 < j \leq i+1).$$

（2）当 $\hat{Z}_{n-1}>0$ 时，若 $\hat{Z}_n=0$，则在 $(t_{n-1},t_n]$ 有 $i+1$ 个顾客离开，而且还出现了闲期,于是在条件 $\{\hat{Z}_n=0\}$ 下 $\{\hat{Y}_{n-1}=i+1\}$ 不是等价于 $\{N'_{t_n}-N'_{t_{n-1}}=i+1\}$，而是等价于 $\{N'_{t_n}-N'_{t_{n-1}}\ge i+1\}$．所以，与上面类似地有

$$P(\hat{Z}_n=0\mid \hat{Z}_{n-1}=i)=P(\hat{Y}_{n-1}=i+1)$$

$$=\int_0^\infty P(N'_{\tau_n}-N'_{\tau_{n-1}}\ge i+1\mid T_n=t)g(t)dt=\int_0^\infty e^{-m\cdot t}\sum_{k=i+1}^\infty\frac{(m\cdot t)^k}{k!}g(t)dt=b_j\ .$$

定理７．９ 对于排队系统 $G/M/1$，在

$$\int tg(t)dt>\frac{1}{m} \qquad （７．３８）$$

时，排队过程在顾客到达的时刻所导出的 Markov 链有形式如

$$p=\frac{1}{1-b}(1,b,\cdots,b^n,\cdots) \qquad （７．３９）$$

的不变分布，其中 b 是方程

$$b=\int_0^\infty e^{-m(1-b)t}g(t)dt \qquad （７．４０）$$

在 $(0,1)$ 中的唯一解．而且此时有 $\mathbf{P}^n\to\mathbf{1}^Tp$ ．

证明 Markov 链 X_{t_n} 显见互通且是非周期的．如果我们证明了它至少存在一个不变分布 p ，那么它一定是正常返的，从而不变分布唯一，而且 $\mathbf{P}^n\to\mathbf{1}^Tp$ ．在（７．３８）条件下时，我们待定常数 b ，使以（７．３９）定义的 p 是不变分布．而不变分布的条件 $p=\mathbf{P}p$ 即

$$b^n=\sum_{k=n-1}^\infty b^k b_{k-(n-1)}=b^{n-1}\sum_{j=0}^\infty b^j b_j\ .$$

利用（７．３７）就得到 b 满足（７．４０）．而在条件（７．３８）下, 可以用分析方法证明（因为较繁，所以略去)上面的方程在开区间 $(0,1)$ 中存在唯一解．

［注］ 在 3．3 段与 3．4 段中，若 $G(x)$ 没有密度，那么，只要用 $dG(t)$ 代替 $g(t)dt$，即只要把普通积分改为 Stieltjes 积分，就使所有的结论保持正确．

3．5　关于 $M/G/\infty$ 系统的注记

$M/G/\infty$ 系统与 $M/G/1$ 系统的唯一不同是，$M/G/\infty$ 系统有 ∞ 条服务线，进入系统的每个顾客可以随机地选取一条服务线．而各服务线的服务时间仍是服从分布函数 $G(t)$ 的独立同分布随机变量，且与输入流彼此独立．为了在数学上简单些，我们仍假定 $G(t)$ 有密度 $g(t)$．

我们来考虑 $M/G/\infty$ 系统的输出(顾客)流．由于有 ∞ 条服务线，进入系统的顾客立刻得到服务，顾客进入后，在经过时间 t 以前离开的概率就是服务时间不大于 t 的概率，即是 $G(t)$．若我们把离开与还在接受服务看成顾客流的随机分流，那么这个分流的概率就是 $p=G(t)$．由 Poisson 过程的非时齐分流定理，我们立刻得到下述结论：

定理７．１０　$M/G/\infty$ 系统的输出流是强度为 $lG(t)$ 的非齐次Poisson流，即对于第 n 个顾客离开系统的时间 t_n，其计数过程 $N_t \overset{\Delta}{=} \sup\{n: t_n \le t\}$ 是强度为 $lG(t)$ 的非时齐 Poisson 过程．

这时服务时间长度 $\tilde{T}_i$ 为独立同分布的任意分布函数 $G(t)$．只要 $m \overset{D}{=} E\tilde{T}_i > 0$ 有限,那么经过并不复杂的推导(参见： A.я.欣钦， 公用事业理论的数学方法，科学出版社 1958 年中译本,第 72 页)，就可得到：

定理７．１１　当 $t\to\infty$ 时，排队过程 X_t 近似于 Poisson 分布 $Poisson_{l\cdot m}$．　】

对于一般排队过程来说，通过定理４．２８用混合的 Erlang 分布，来近似排队系统中的服务分布或顾客到达的间隔分布的方法，是一个简化排队系统的重要方法，通常称为位相方法．例如，由定理 4.28，排队系统中的服务时间的近似分布为 $F_h(x)$，它对应于顾客要以概率 p_n 连续地完成 n 个独立的串联的分布为 $Exp_{\frac{1}{h}}$ 的服务后(称为 n 个位相)，才离开系统．这种思路可以用来有限地简化排队系统的一些统计指标的近似计算，如不变分布，队伍长度的分布等等．

[注] 处理一般非 Markov 排队过程 X_t 的常用的方法中，还有一种加入辅助变量使之成为高一维的 Markov 过程的辅助过程方法．例如，对于 $M/G/1$ 系统的排队过程 X_t，记 $j_t = t$ 时刻在接受服务的顾客已接受服务的时间，那么 (X_t, j_t) 是取值于 $Z^+\times R^+$ 的 Markov 过程，其中 Z^+, R^+ 分别为非负整数集，非负实数集．类似地，对于 $G/M/1$ 系统的

排队过程 X_t，只要记 $y_t=$“自 t 时刻至下一个顾客到达的等待时间”，则 (X_t, y_t) 也是取值于 $Z^+ \times R^+$ 的 Markov 过程．这里的 j_t, y_t 都是辅助随机过程．辅助随机过程引入后，就可以使用 Kolmogorov 方程和 Master 方程，以及不变分布等工具．近年来我国有些学者，注意到了多数排队系统的排队过程在某些特定的随机时刻上具有 Markov 性，并称这种在一些特殊时刻上具有 Markov 性的随机过程为 Markov 骨架过程．他们利用首达时刻分析方法，得到了 Markov 骨架过程的转移特征满足的积分方程，由此可以得到描述排队过程的许多特征量．在理论上为研究 $GI/G/1$ 等系统，提供了不同于在辅助过程方法的一个途径．

*4. 半 Markov 过程

4.1 半 Markov 过程

1． 半 Markov 过程是时间连续的 Markov 链的自然推广，它并不假定呆在一个状态的时间服从指数分布． 因此，它更多地出现在许多常见的排队过程中．它的确切定义为

定义 7．12 设 x_t 是一个最多只取可数个状态的连续时间的随机过程， 其跳跃由某个连续时间的 Markov 链的嵌入链的转移矩阵决定，但是它呆在状态 i 的时间未必服从指数分布，假定在状态 i 呆的时间是与离开 i 时进入的状态 j 也有关的某个分布 $F_{ij}(t),(i \neq j)$．再假定各次跳跃是独立的，呆在不同状态的时间是独立的，且它们之间也相互独立． 这样的随机过程 x_t 称为半 Markov 过程．

2． 半 Markov 过程有与连续时间的 Markov 链类似的优点，就是容易对它作随机模拟．我们以 N 个状态的半 Markov 过程为例，给出它的轨道模拟步骤如下：

(1). 模拟初始分布 $m_0=(m_0(1),\cdots,m_0(N))$： 对于 $1 \le i \le N$，以概率 $m_0(i)$ 取 i；

(2). 作分布 F_{ij} 的随机数 T_{ij}，让轨道在状态 i 停留时间 T_{ij} 后，以概率 p_{ij} $(\sum_{j \neq i} p_{ij}=1)$ 跳到 $j(\neq i)$；

(3). 以 j 代替 i 后重复步骤(2)..

这个 $\mathbf{P}=(p_{ij}),(p_{ii}=0)$ 对应的 Markov 链，称为半 Markov 过程的嵌入链，也称 $\mathbf{P}$ 为半 Markov 过程的嵌入转移矩阵．

3． 半 Markov 过程在转移前平均停留时间可以表达如下：

设从 i 出发，到转移前在 i 停留的时间为 t_i．由假定，在转移到 j 的条件下，t_i 的条件分布为 F_{ij}，而从 i 转移为 j 的概率为 p_{ij}，于是 t_i 的分布函数为

$$P(t_i \le t)=\sum_{j \neq i} p_{ij} F_{ij}(t).$$

4.2 半 Markov 过程的渐近性质

令 t_{ii} 为半 Markov 过程两次连续进入 i 间的一个循环的时间间隔， 那么半 Markov 过程从开始停留在

i，一直到离开 i 后又首次回到 i，就形成一个循环， 所有的循环连起来就成为一个交错更新过程．对这个交错更新过程用极限定理，便得到半 Markov 过程的状态转移的渐近性质．

定理7．13 如果半 Markov 链的嵌入转移矩阵 $\mathbf{P}$ 不可约， 且 t_{ii} 的分布为非格点分布，又 $Et_{ii} < \infty$．那么

$$P(x_t = j \mid x_0 = i) \xrightarrow{t\to\infty} p_j \overset{\Delta}{=} \frac{Et_j}{Et_{jj}},$$

$$\frac{[0,t]\text{中过程停留在}j\text{的时间}}{t} \xrightarrow{t\to\infty} \frac{Et_j}{Et_{jj}} = p_j,$$

它们都与半 Markov 链的初值无关．】

若嵌入链 $\mathbf{P}$ 正常返，互通，即 $\mathbf{P}^n \to \tilde{p}$ ．又若 t_{ii} 为非格点分布，且 $Et_{ii} < \infty$．回忆 Markov 链的情形．那时有 $p_j = C\dfrac{\tilde{p}_j}{q_j} = C\tilde{p}_j Et_j$． 可以猜想这个关系对于半 Markov 过程仍然正确．即我们有下述定理

定理7．14 若半 Markov 过程的嵌入链 $\mathbf{P}$ 正常返，互通，即 $\mathbf{P}^n \to \tilde{p}$ ．又若 t_{ii} 为非格点分布，且 $Et_{ii} < \infty$，则

$$p_j (\overset{\Delta}{=} \frac{Et_j}{Et_{jj}}) = \frac{\tilde{p}_j Et_j}{\sum_k \tilde{p}_k Et_k},$$

它等价于

$$Et_{jj} = \frac{\sum_k \tilde{p}_k Et_k}{\tilde{p}_j}.$$

证明 令 $t_j^{(k)}$ 为第 k 次在 j 停留的时间长度，$N_j^{(n)}$ 为前 n 次转移中取 j 的次数， $p_j^{(n)}$ 为前 n 次转移中取 j 的时间的比例．用 $p_j = \lim_{n\to\infty} p_j^{(n)}$ 及强大数定律于

$$p_j^{(n)} = \frac{\sum_{k=1}^{N_j^{(n)}} t_j^{(k)}}{\sum_i \sum_{k=1}^{N_i^{(n)}} t_i^{(k)}} = \frac{\frac{N_j^{(n)}}{n} \cdot \frac{1}{N_j^{(n)}} \sum_{k=1}^{N_j^{(n)}} t_j^{(k)}}{\sum_i \frac{N_i^{(n)}}{n} \cdot \frac{1}{N_i^{(n)}} \sum_{k=1}^{N_i^{(n)}} t_i^{(k)}}$$

$$\xrightarrow{n\to\infty} \frac{\tilde{p}_j Et_j}{\sum_i \tilde{p}_i Et_i},$$

便得定理. 】

利用交替更新定理, 我们还可以证明

命题7。16

$$P(x_t = j, t\text{后首次转移在}t+s\text{后},\text{转移到}k \mid x_0 = i) = \frac{p_{jk}\int_s^\infty (1-F_{jk}(u))du}{Et_{jj}}.$$

$$P(x_t = j, t\text{后首次转移在}t+s\text{后} \mid x_0 = i) = \frac{\int_s^\infty P(t_j > u)du}{Et_{jj}}.$$

证明　记服从分布 F_{ij} 的随机变量为 t_{ij}. 在两次连续进入 i 间的一个循环的时间间隔中, 满足条件

$$\{x_t = j, t\text{后首次转移在}t+s\text{后},\text{转移到}k\}$$

的时间, 与不满足该条件的时间, 分别组成交替更新间隔. 而在一个 i-循环中满足此条件的时间的期望为

$$p_{jk}E[(t_{jk}-s)^+] = p_{jk}\int_0^\infty P((t_{jk}-s)^+ > u)du$$

$$= p_{jk}\int_0^\infty P(t_{jk} > u+s)du = p_{jk}\int_s^\infty P(t_{jk} > u)du.$$

用交替更新定理便得第一个等式. 而第二个等式乃是第一个等式的推论.

* 5. 有限位相型分布(PH-分布)

5. 1　背景

在排队系统的理论中, 常出现所谓有限位相型分布(简称为 PH-分布). 这类分布是指数分布在矩阵意义下的推广. 这一类分布包括了混合 Erlang 分布，因此，可以用来近似任意有密度的正随机变量的分布. 所以, PH 分布类有广泛的代表性. 即在$[0,\infty)$上取值的随机变量的分布都可以用 PH 分布近似.

直观地, 假定给定了一个具有一个吸收状态的有限状态的连续时间的 Markov 链. 如果在此吸收态设置一个观测点, 测量此 Markov 链在被吸收的时间, 即首次到达此吸收点的时刻 t . 这个随机时间 (它是停时) 的分布, 就称为 有限位相型分布(PH-分布), t 就称为 PH 随机变量.

不同的连续时间的 Markov 链, 可能生成相同的有限位相型分布.

一个相反的问题是: 知道了这个有限位相分布(PH-分布), 能在多大程度上知道此连续时间的 Markov 链的转移矩阵，或它的转移速率矩阵呢 ? 如果回答为肯定, 那么这就提供了一种测量转移参数的手段.

下段中我们把上面的定义, 叙述得更为数学化一些.

5. 2　PH 分布 (有限位相型分布)

定义７．１７ 设状态空间为$\{0,1,\cdots,N\}$的连续时间的 Markov 链x_t，其转移速率矩阵为

$\mathbf{Q}=\begin{pmatrix} 0 & 0 \\ \mathbf{q}^{\mathbf{T}} & \tilde{\mathbf{Q}} \end{pmatrix}$,（$\mathbf{Q}$的行和为 1）. 再假定$\tilde{\mathbf{Q}}$是一个对角线上为恒负的可逆矩阵. 此时 0 为唯一的吸收态. 我们将x_t被 0 吸收的时刻t 的分布称为 PH 分布.

下面我们将 PH 分布写为参数形式. 设x_t 的初始分布为

$$P(x_0 = j) = a_j, (0 \le j \le N, \sum_{j=0}^{N} a_j = 1).$$

仿照离散时间 Markov 情形，我们记禁忌到过 0 的转移概率为${}_{\{0\}}p_{ij}(t)$，那么类似地有

$$({}_{\{0\}}p_{ij}(t)) = e^{-\tilde{\mathbf{Q}}t}.$$

于是 PH 分布可以写为

$$F(t) \overset{D}{=} P(t \le t) = 1 - P(t > t)$$

$$= 1 - \sum_{i,j=1}^{N} P(x_t = j, x_s \ne 0 (0 < s < t) \mid x_0 = i) P(x_0 = i)$$

$$= 1 - \sum_{i,j=1}^{N} {}_{\{0\}}P_{ij}(t) P(x_0 = i)$$

$$= 1 - (a_1, \cdots, a_N) e^{\tilde{\mathbf{Q}}t} \mathbf{1}^T.$$

由此得到

定理７.１８ PH 分布的分布函数总可以表示为$1-(a_1,\cdots,a_N)e^{\tilde{\mathbf{Q}}t}\mathbf{1}^T$，其中

$\tilde{\mathbf{Q}} = \left(\tilde{q}_{ij}\right)$，它们满足

$$a_i \ge 0, \sum_{i=1}^{N} a_i < 1, \tilde{q}_{ii} < 0, \tilde{q}_{ij} \ge 0, (i \ne j).$$

所以 PH 分布只依赖参数$a \overset{\Delta}{=} (a_1, \cdots, a_N)$与$\tilde{\mathbf{Q}}$.

定义７．１９ 定理７．１８中的参数组$(a, \tilde{\mathbf{Q}})$，称为 PH 分布的一个参数表示. 不同的参数组可能对应与相同的 PH 分布. 因此，同一个 PH 分布可以有多个参数表示. N 最小的那组参数表示，就称为此 PH 分布的最小表示.

例７．２０ 参数为l 的指数分布 $\exp_l$ 是 PH 分布，且 它的最小表示为：$a \overset{\Delta}{=} 1$，$\tilde{\mathbf{Q}} = -l$.

证明 记以$\{0,1\}$为状态空间，并以$\mathbf{Q}=\begin{pmatrix}0 & 0\\ l & -l\end{pmatrix}$为转移速率矩阵的连续时间的 Markov 链为$x_t$，则由上面的结论可知，$x_t$首达 0 的时刻$t$ 的分布为$P(t>t)=e^{-lt}$. 这正说明了$t\sim\exp_l$.

命题７．２１ Erlang 分布 （$\Gamma(k,l)$.分布）是 PH 分布.

证明 在$k=1$时，$\Gamma(k,l)$分布就是指数分布. 为了得到一些直观的认识，我们先看$k=2$情形. 我们证明$\Gamma(2,l)$分布是以：$a=(1,0)$，$\tilde{\mathbf{Q}}=\begin{pmatrix}-l & l\\ 0 & -l\end{pmatrix}$为最小表示的 PH 分布. 为此，我们记以$\{0,1,2\}$为状态空间，初值$x_0=1$，并以$\mathbf{Q}=\begin{pmatrix}0 & & \\ 0 & -l & l\\ l & & -l\end{pmatrix}$为转移速率矩阵的连续时间的 Markov 链为$x_t$. 利用

$$\begin{pmatrix}-1 & 1\\ & -1\end{pmatrix}^n=\begin{pmatrix}(-1)^n & (-1)^{n-1}n\\ & (-1)^n\end{pmatrix},$$

我们得到

$$e^{\tilde{\mathbf{Q}}\,t}=e^{\begin{pmatrix}-1 & 1\\ & -1\end{pmatrix}lt}=\begin{pmatrix}e^{-lt} & \sum_{n=0}^{\infty}(-1)^{n-i}n\dfrac{(l\cdot t)^n}{n!}\\ & e^{-lt}\end{pmatrix}=\begin{pmatrix}e^{-lt} & lte^{-lt}\\ & e^{-lt}\end{pmatrix}$$

于是由定理７.１８知道x_t首达 0 的时刻t 的分布为

$$P(t>t)=(1,0)e^{\tilde{Q}\,t}\begin{pmatrix}1\\1\end{pmatrix}=(1+lt)e^{-lt}.$$

由此得到t 的分布密度为 $l^2te^{-lt}I_{[0,\infty)}(t)$，即为$\Gamma(2,l)$分布. 从而$\Gamma(2,l)$分布为 PH 分布.

对于一般的k.我们记以$\{0,1,\cdots,k\}$为状态空间，初值$x_0=1$，并以

$$\mathbf{Q}=\begin{pmatrix}0 & & & & \\ & -l & l & & \\ & & -l & l & \\ & & & \ddots & \ddots\\ l & & & & -l\end{pmatrix}$$

为转移速率矩阵的连续时间的 Markov 链为x_t. 利用

$$\begin{pmatrix} -1 & 1 & & \\ & -1 & 1 & \\ & & \ddots & \ddots \\ & & & -1 \end{pmatrix}^n$$

$$= \begin{pmatrix} (-1)^n & (-1)^{n-1}n & (-1)^{n-2}n(n-1) & \cdots & (-1)^{n-k}n(n-1)\cdots(n-k+1) \\ & (-1)^n & (-1)^{n-1}n & (-1)^{n-2}n(n-1) & \cdots \\ & & \ddots & \ddots & \ddots \\ & & & \ddots & (-1)^{n-1}n(n-1) \\ & & & & (-1)^n \end{pmatrix},$$

我们得到

$$e^{\tilde{\mathbf{Q}}\,t} = e^{\begin{pmatrix} -1 & 1 & & \\ & -1 & 1 & \\ & & \ddots & \ddots \\ & & & -1 \end{pmatrix} lt} = e^{-l\cdot t}\begin{pmatrix} 1 & lt & (lt)^2 & \cdots & (lt)^{k-1} \\ & 1 & lt & (lt)^2 & \cdots \\ & & \ddots & \ddots & \ddots \\ & & & 1 & lt \\ & & & & 1 \end{pmatrix}.$$

由定理７.１８知道x_t **首达** 0 **的时刻**t **的分布为**

$$P(t > t) = (1,0,\cdots 0)e^{\tilde{Q}\,t}\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = (1 + lt + \frac{(lt)^2}{2!} + \cdots + \frac{(lt)^{k-1}}{(k-1)!})e^{-lt}.$$

即t **的分布密度为** $l\,\frac{(lt)^{k-1}}{(k-1)!}e^{-lt}I_{[0,\infty)}(t)$，**就是**$\Gamma(k, l)$**分布. 从而**$\Gamma(k, l)$**分布为** PH **分布.**

命题７．２２　PH **随机变量的混合仍为** PH **随机变量, 即: 若** $F(x), G(x)$ **均为** PH **分布, 则** $lF(x) + (1-l)G(x), (0 < l < 1)$，**也是** PH **分布.**

（证明提示　设它们分别有表示$(a^{(1)}, \tilde{\mathbf{Q}}^{(1)})$ **与** $(a^{(2)}, \tilde{\mathbf{Q}}^{(2)})$ **，那么由直观可以看出，**$((l\,a^{(1)}, (1-l)\,a^{(2)}), \begin{pmatrix} \tilde{\mathbf{Q}}^{(1)} & \\ & \tilde{\mathbf{Q}}^{(2)} \end{pmatrix})$ **就是** $lF(x) + (1-l)G(x)$ **的一个表示).**

例７．２３　（超指数分布 －H_m**分布）**

$$F(t) = \sum_{i=1}^{m} a_i(1 - e^{-l_i t}),$$

其表示为$(a, \tilde{\mathbf{Q}}), a = (a_1, \cdots, a_m)$，$\tilde{\mathbf{Q}}$**是对角元素为**$-l_1, \cdots, -l_m$**的对角矩阵.**

与超指数分布有关的还有$Coxian-2$**型分布．假定随机变量**X_1, X_2 **相互独立，分布服从分布**

Exp_{l_1}, Exp_{l_2}. 而随机变量x是混合型的，它以概率 p 取 X_1+X_2，而以概率$1-p$ 取 X_1，则随机变量x 的分布称为$Coxian-2$型分布.

定理7．23 PH 分布是常数与连续型随机变量的混合分布

证明 显见 $P(t=0)=a_0$. 而且在$a_0=0$时，t 具有分布密度(利用$\mathbf{Q}$ 的行和为 1，即 $\tilde{\mathbf{Q}}\mathbf{1}=\mathbf{q}$)

$$f_+(t)=(a_1,\cdots,a_N)e^{-\tilde{\mathbf{Q}}t}\mathbf{q}^T.$$

然而，当$a_0>0$时，$\int_0^\infty f_+(t)dt=\sum_{k=1}^N a_k<1$，此时$\frac{1}{1-a_0}f_+(t)$才是分布密度. 于是 PH 随机变量$t$的分布是如下的一个混合分布:

以a_0的概率取常值 0，而以 $1-a_0$的概率具有分布密度$\frac{1}{1-a_0}f_+(t)$.

在实际情形中，常出现$a_0=0$情形，此时 PH 分布是有密度的. 注意此“有密度部分”看起来像“矩阵型指数分布”. 这是指数分布的一种推广.

命题7.24 （PH 分布的矩母函数）

若随机变量t 服从 PH 分布，则它的矩母函数为

$$M(s)\overset{\Delta}{=}Ee^{st}=a_0e^0+(1-a_0)\int_0^\infty e^{st}\frac{1}{1-a_0}f_+(t)dt$$

$$=a_0+(a_1,\cdots,a_N)\ (sI-\tilde{\mathbf{Q}})^{-1}\mathbf{q}^T.$$

而其k 阶矩为

$$Et^k=\frac{d^kM(s)}{ds^k}.$$

5. 3 **离散位相型分布**

与 PH 分布的不同处是，这里用离散时间的 Markov 链x_n来代替前面的连续时间 Markov 链x_t.

定义7．25 设x_n的状态空间为{0,1,…,N}，初分布为

$$P(x_0=j)=a_j,(0\le j\le N,\sum_{j=0}^N a_j=1),$$

转移矩阵为

$$\mathbf{P}=\begin{pmatrix}1 & 0\\ \mathbf{p}^{\mathrm{T}} & \tilde{\mathbf{P}}\end{pmatrix}，其中\mathbf{I}-\tilde{\mathbf{P}}为可逆.$$

x_n 被状态 0 吸收的时刻 t ，是一个取非负整值的随机变量，称为离散的 PH 随机变量．它的分布称为离散 PH 分布． 此分布依赖于参数组 $(a,\tilde{\boldsymbol{P}}), a=(a_0,\cdots,a_N)$．

同样地，不同的参数组可能对应与同一个离散 PH 分布. 因此离散 PH 分布也可能有不同的参数表示. N 最小的那组参数, 就称为此 PH 分布的最小表示.

命题７.２６ （离散 PH 分布的参数表示）

仿 PH 分布情形, 对于离散 PH 分布, 易得

$$P(t=k)=\begin{cases} a_0 & (k=0) \\ (a_1,\cdots,a_N)\,\tilde{\mathbf{P}}^{k-1}\,\mathbf{p}^T & (k=1,2,\cdots) \end{cases}.$$

这是一个 "矩阵型几何分布". 由此可以直接得到它的母函数

$$P(z)=a_0+z(a_1,\cdots,a_N)(\mathbf{I}-z\tilde{\mathbf{P}})^{-1}\mathbf{p}^T$$

和各阶矩.

5.4 PH 分布族的封闭性

经过仔细的分析运算后可证明, PH 分布类在多种运算下具有封闭性, 即有:

（１） 独立 PH 随机变量的和仍为 PH 随机变量.

(证明提示 设它们分别有表示 $(a^{(1)},\tilde{\mathbf{Q}}^{(1)})$ 与 $(a^{(2)},\tilde{\mathbf{Q}}^{(2)})$，那么可以证明 $\mathrm{a}=(a_0^{(1)}\ \ \mathrm{a}^{(2)})$，

$$\tilde{\mathbf{Q}}=\begin{pmatrix}\tilde{\mathbf{Q}}^{(1)} & 0\\ \mathbf{q}^{(1)\mathbf{T}}\mathrm{a}^{(2)} & \tilde{\mathbf{Q}}^{(2)}\end{pmatrix}$$

就是此两个随机变量的和的一个 PH 型的表示)．

（２） 若 x,h 都是 PH 随机变量, 则它们的最大值 $x\vee h$, 最小值 $x\wedge h$, 都是 PH 随机变量.

（３）. 两个 PH 分布的混合分布, 仍是 PH 分布.

（４） 若 $\{t_k\}$ 为独立同 PH 分布, h 为与它们独立的几何随机变量, 那么 $\sum_1^h t_i$ 也服从 PH 分布.

［注１］ 还可以考虑如下的广义 PH 型分布 $F(t)=1-(a_1,\cdots,a_N)e^{\tilde{\mathbf{Q}}\,h(t)}\mathbf{1}^T$，其中 $h(t)$ 为 $(-\infty,\infty)$ 上定义的偶函数，满足 $h(0)=0, h(t)$ 在 $[0,\infty)$ 上非降． 有人用广义 PH 型分布的波形设计滤波器。

［注２］ 关于排队系统 $G/E_k/1$，$E_k/G/1$，$G/PH/1$ 与 $PH/G/1$ 的讨论可参见；

徐光辉，随机服务系统，第２版，科学出版社，１９８８．

徐光辉，袁学明， 排队论，现代数学手册，随机数学卷，第１５篇，华中科技大学出版社，２０００．

习题 7

1. 设 $\{x_n\},\{h_n\}$ 是相互独立的独立同分布随机变量列， $x_1\sim\begin{pmatrix}1 & 2\\ 1-a & a\end{pmatrix}$, $h_1\sim\begin{pmatrix}1 & 2\\ 1-b & b\end{pmatrix}$.

假定只有排队系统一条服务线，顾客都在整数分钟的时间到达服务线，且同一时间只会有一个顾客到达.而第n个顾客到达的时刻为$t_n \overset{\Delta}{=} x_1 + \cdots + x_n$. 再假定也只在整数分钟的时间才接受服务,先到先服务.第n 个顾客需要服务的时间为h_n. 记时刻n的队伍长度为X_n （包括等待的顾客和在接受服务且未完的顾客， 称为离散时间的排队过程）. 证明X_n是Markov链，并求其平稳分布.

2. 发廊有N个美发师，且只能容纳$M(>N)$个顾客). 假定第$M+1$个顾客会自动离去. 设顾客以强度为λ的指数流到达，而做头发所需时间服从参数为μ的指数分布，而且与到达的情形相互独立.求在时刻t在发廊中顾客数X_t的平稳分布.

3. 在$M/M/2$系统中，设顾客按参数为λ的指数流到达，而服务时间是与之独立的参数为μ的独立指数分布，且$l < 2m$. 如果改用一个比原来的两个服务员快一倍的一个服务员(即改用$M/M/1$系统，但是服务时间的分布为$\exp_{2m}$). 请比较这两种服务系统在稳定情形下的平均排队的队伍长度，并判断其优劣.

4. 在$M/M/1$中若每当队伍长度大于N时就增加一条服务线，而当队伍长度$\leq N$时就撤掉新增加的那条服务线.求排队队伍长度X_t的平稳分布(队伍长度是指在系统中顾客的数目)

5 对于 $M/M/1$排队系统，记在一个服务间隔中到达的顾客数为n的概率为h_n. 用向前方程导出a_n满足的差分方程，从而证明在一个服务间隔中到达的顾客数的分布为参数为$\dfrac{l}{l+m}$的几何分布.

6.（带随机消失的$M/M/1$）

在下列情形下分别求在时刻t系统中顾客数X_t的平稳分布：

(1).新来的顾客看到原来的$X_t > k$ （k>1)则以概率p留下，而以概率$1-p$离去.

(2).新来的顾客看到原系统中的人数n超过k则离去，而在$n \leq k$时则以概率$p = \dfrac{k-n}{k(n+1)}$留下，而以概率$1-p$离去.

(3). 把(2)中的p改为$\dfrac{1}{1+n}$.

7. 出租车站以指数流λ来车，而顾客则以指数流μ独立的到站.假定一辆车只载一个客人,且顾客到站时如看到已有两人在等候时就自动离开. 设$l < m$. 问在稳定情形下，平均有客人在等候汽车和平均有多少出租车在等候客人 ？

8. 分别在$GI/M/N$和$GI/M/\infty$系统中，把输入流的计数过程由Poisson过程改为复合Poisson过程，说明此时在时刻t在系统中的顾客数X_t仍为Markov链，并写出它的转移速率阵Q，再讨论什么时

候有平稳分布,这平稳分布是什么.

9. 设顾客按参数为 l 的指数流到达某个服务机构，每人必须相继地接受独立的，服务时间分别服从参数为 m_1 和 m_2 的指数分布的两次服务. 假定只在两个服务台都有空时顾客才接受服务，否则就离开.求顾客能接受服务的概率.

10. 某实验室有 n 台相同的仪器，它们独立地处在工作状态或调试状态. 工作的仪器每台耗电 a 千瓦，调试的仪器每台耗电 b 千瓦. 在时刻 t 工作着的仪器于 $(t,t+h)$ 内转为调试的概率为 $l\cdot h+o(h)$，而在时刻 t 调试的仪器于 $(t,t+h)$ 内重新工作的概率为 $m\cdot h+o(h)$. 求此实验室的平均耗电.

11. 系统由 n 个独立地工作的相同元件并联而成，元件的寿命服从参数为 l 的指数分布. 另有 m 个同类的备用元件. 又一个失效就立刻换上. 备用件用完后就不再更换. 证明此系统能工作的时间的分布函数为 $\int_0^t (1-e^{-l(t-s)})^{m-1}\frac{(ml)^{n+1}}{n!}s^n e^{-ml\cdot s}ds$.

12. 假定顾客按强度为 l 的指数流到达服务台，如果他看见有 n 人在排队，则他以概率 $\frac{1}{2^n}$ 排队，而以概率 $1-\frac{1}{2^n}$ 离去. 假定服务台的服务时间服从与之独立的 $\exp_m$. 将时刻 t 在服务台前排队的人数(包括正在接受服务的人)记为 X_t，再记其分布为 $p_n(t)=P(X_t=n)$. 证明它满足方程

$$p'_n(t)=mp_{n+1}(t)-(\frac{l}{2^n}+m)p_n(t)+\frac{l}{2^{n-1}}p_{n-1}(t)\quad (n\geq 1),$$

$$p'_0(t)=mp_1(t)-lp_0(t).$$

进而证明它有平稳分布 $p_n=p_0\frac{1}{2^{\frac{n(n-1)}{2}}}\left(\frac{l}{m}\right)^n$ ($n\geq 0$).

龚光鲁，钱敏平著 应用随机过程教程及其在算法与智能计算中的应用

清华大学出版社,2003

第 8 章 Monte Carlo 与 Markov Chain Monte Carlo (MCMC) 方法

在许多很复杂的统计问题中,有时很难直接对各种统计方法进行理论分析．为了评估它们的优劣,常见的实用办法是做随机模拟：即设法按问题的要求与条件去构造出一系列的模拟样本,用它们的样本频率代替对应的概率作统计分析与推断，观察由这些摸拟样品所作出的推断的正确率．因为在概率论初期发展时，随机模拟的原型常常来自博采，于是人们就以博采之都 Monte Carlo 作为随机模拟方法的别称．久而久之，Monte Carlo 方法作为名称倒比随机模拟方法更为广泛地被常用了．相仿地，人们还把组合计算中的某些随机模拟方法，称为 Las Vegas 方法，这是以美国的博采城 Las Vegas 命名的．

一般随机模拟方法的长处是,计算的复杂度不依赖于计算空间的维数. 因此, 在计算非常高维数的积分（或多指标的求和）时, Monte Carlo 方法比通常的计算方法有明显的优越性. 而很多的 Monte Carlo 计算问题, 可归结为计算积分或庞大和数的问题.

1. 计算积分的 Monte Carlo 方法与采样量估计

通过构造独立同分布随机数，计算积分的 Monte Carlo 方法，也称为静态 Monte Carlo 方法，其思想可以在本节中,通过估计最简单的积分 $\int_a^b f(x)dx$ 得到阐明．对于高维积分，其思路与一维积分是一样的．另一方面，在甚高维情形，因为计算量太大，用静态 Monte Carlo 方法处理速度太慢． 我们在第 2 节中通过构造 Markov 链的极限不变分布，来模拟计算积分的方法（称为动态 Monte Carlo 方法， Markov 链 Monte Carlo 方法，MCMC)，将更为适用．

1.1 用频率估计概率来计算积分的 Monte Carlo 方法

假定 $0 \le f(x) \le M$ ．那么由积分的面积含义有

$$\int_a^b f(x)dx = |S|, \qquad (\text{其中} |S| \text{ 为 } S \overset{\Delta}{=} \{(x,y): a \le x \le b, f(x) \ge y\} \text{ 的面积 }).$$

考虑平面区域 $\Omega \overset{\Delta}{=} [a,b]\times[0,M]$ 上的均匀随机变量 x，则

$$p \overset{\Delta}{=} P(x \in S) = \frac{1}{(b-a)M} \int_a^b f(x)dx .$$

对于 N 个独立的 Ω -均匀随机数 $x_i, (i \le N)$，记

$$N_S = \{x_1, \cdots, x_N\} \text{ 中落在 } S \text{ 中的频数，}$$

于是，利用大数定律便知

$$\hat{I} \overset{\Delta}{=} (b-a)M \frac{N_S}{N} \qquad (8.1)$$

是积分 $I \overset{\Delta}{=} \int_a^b f(x)dx$ 的相合估计，即对于任意的 $e>0$，当 $N\to\infty$ 时，有

$$P(|(b-a)M\frac{N_S}{N}-\int_a^b f(x)dx|>e)\to 0 .$$

又由于 N_S 服从参数为 (p,N) 的二项分布，所以我们有 $E\hat{I}=\int_a^b f(x)dx$，即 $\hat{I}$ 是积分 $\int_a^b f(x)dx$ 的无偏估计．此估计的方差为

$$Var\,\hat{I}=(b-a)^2M^2\frac{p(1-p)}{N}=(b-a)^2M^2\frac{\frac{I}{(b-a)M}(1-\frac{I}{(b-a)M})}{N}$$

$$=\frac{1}{N}[(b-a)M-I]I \quad =O(\frac{1}{N}) .$$

又因为方差代表平均平方误差，所以我们就说，积分的估计 $\hat{I}$ 的误差为 $O(\frac{1}{\sqrt{N}})$．

频率法所需采样量的估计

我们知道统计估计是以区间估计来给出概率误差的，即对于允许误差 e，及允许以 d 的失败概率，来确定样本量 N（依赖于 e,d），使

$$P(|\hat{I}-\int_a^b f(x)dx|>e)<d .$$

1　用 Chebyshev 不等式估计样本量

最粗略的确定样本量 N 的方法，是通过 Chebyshev 不等式来估计．即由

$$P(|\hat{I}-\int_a^b f(x)dx|>e)=P(|\hat{I}-E\hat{I}|>e)\le\frac{Var\,\hat{I}}{e^2},$$

要求 $\frac{Var\,\hat{I}}{e^2}=\frac{(b-a)^2M^2\frac{p(1-p)}{N}}{e^2}<d$．考虑到 $p(1-p)\le\frac{1}{4}$，我们要确定 N，只需满足 $\frac{(b-a)^2M^2\frac{1}{4N}}{e^2}<d$ 就够了，也就是说只要 $N>\frac{(b-a)^2M^2}{4de^2}$ 即可．但是这个采样量的估计过于粗糙.

2　用中心极限定理的近似估计样本量

在 N 大时(这总假定满足)，用中心极限定理也可以得到采样量 N 的估计．中心极限定理断言，$\hat{I}=(b-a)M\frac{N_S}{N}$ 的近似分布为正态分布 $N(\int_a^b f(x)dx, Var\,\hat{I})$．令 f_d 满足

$$\frac{1}{\sqrt{2p}}\int_{f_d}^{\infty} e^{-\frac{x^2}{2}}dx = d \text{ ,}\quad (\text{ 即 } \frac{1}{\sqrt{2p}}\int_{-f_{\frac{d}{2}}}^{f_{\frac{d}{2}}} e^{-\frac{x^2}{2}}dx = 1-d \quad).$$

于是有

$$P(|\hat{I}-\int_a^b f(x)dx| \ge f_{\frac{d}{2}}\sqrt{Var\hat{I}}) \le d \text{ .}$$

选取采样量 N 使 $f_{\frac{d}{2}}\sqrt{Var\hat{I}} \le e$，即 $(b-a)^2M^2\frac{p(1-p)}{N} \le \frac{e^2}{f_{\frac{d}{2}}{}^2}$，则就有

$$P(|\hat{I}-\int_a^b f(x)dx| \ge e) \le d \text{ .}$$

所以只需取

$$N > \frac{(b-a)^2M^2}{4e^2}f_{\frac{d}{2}}{}^2.$$

3　采样量的估计的比较

我们把用 Chebyshev 不等式得到的采样量的估计与用中心极限定理得到的采样量的估计作比较，可以看到只要 $df_{\frac{d}{2}}{}^2 << 1$，用中心极限定理得到的采样量的估计，要比用 Chebyshev 不等式得到的采样量的估计小得多.

［注］　还可以应用概率理论中的大偏差速率函数来求得采样量的估计. 可以证明由大偏差速率函数能得到采样量的估计

$$N > \frac{-\ln d}{2e^2} \text{ .}$$

当 d 小时，$-\ln d$ 趋于 ∞ 的速度要比 $f_{\frac{d}{2}}{}^2$ 趋于 ∞ 的速度快. 所以，以用中心极限定理得到的采样量的估计为最好.

１．2　用样本函数的平均值估计积分的 Monte Carlo 方法　—期望法

期望法的核心思想是把积分看成某个随机变量的期望. 最自然的是看成 $[a,b]$ 上均匀随机变量的期望. 设 $h \sim U[a,b]$（$[a,b]$ 上的均匀分布），则

$$I \overset{\Delta}{=} \int_a^b f(x)dx = (b-a)Ef(h) \text{ .} \qquad (8.2)$$

于是对于 N 个独立的 $[a,b]$ 上的均匀随机数，可以用矩估计

$$\overset{\wedge}{\hat{I}} = (b-a)\frac{f(h_1)+\cdots+f(h_N)}{N} \qquad (8.3)$$

作为 $I=\int_a^b f(x)dx$ 的估计．显然它也是无偏估计．而且

$$Var(\hat{\hat{I}})=(b-a)^2\frac{Var[f(h)]}{N}=(b-a)^2\frac{1}{N}\left[Ef(h)^2-[Ef(h)]^2\right]$$

$$=(b-a)^2\frac{1}{N}\int_a^b f(x)^2\frac{dx}{b-a}-\frac{1}{N}I^2\le\frac{1}{N}[(b-a)MI-I^2=Var(\hat{I}).$$

可见期望法比频率法更有效．但是期望法多算了 N 次函数值，这是需要花费计算时间的．所以在全面考虑有效性与计算时间的得失时，有时人们也愿意采用频率法．

1．3　减少方差的技术

用 Monte Carlo 方法计算积分 $\int_a^b f(x)dx$ 时，未必一定要使用均匀随机数．事实上，从 $[a,b]$ 上取值的任意一种随机数出发，都可以得到 $\int_a^b f(x)dx$ 的估计量．而且在 $f(x)\ge 0$ 时显见，值 $f(x)$ 大的 x 对于积分 $\int_a^b f(x)dx$ 有更大的贡献．由此得到直观启发，所用的随机数的分布密度的形状越象 $f(x)$，则就越合理．　这个思想就是重要度采样法．

1.　$g-$采样法

假定分布密度 $g(x)$ 在 $f(x)$ 非零处恒正，则积分 $I=\int_a^b f(x)dx=\int_a^b\frac{f(x)}{g(x)}g(x)dx$．对于密度为 $g(x)$ 的“$g-$随机数” V，我们有

$$I=E[\frac{f(V)}{g(V)}]. \qquad (8.\ 4)$$

于是对于 N 个独立的 $g-$随机数 $V_1,\cdots,V_N$，关于积分 $I=\int_a^b f(x)dx$ 可取估计

$$\hat{I}^{(g)}\overset{\Delta}{=}\frac{1}{N}[\frac{f(V_1)}{g(V_1)}]+\cdots+\frac{f(V_N)}{g(V_N)}]. \qquad (8.\ 5)$$

显见它也是无偏的相合估计．　利用 Schwartz 不等式得到

$$\int_a^b[\frac{f(x)}{g(x)}]^2g(x)dx=\int_a^b g(x)dx\int_a^b[\frac{f(x)}{g(x)}]^2g(x)dx$$

$$\geq [\int_a^b \frac{f(x)}{\sqrt{g(x)}}\sqrt{g(x)}dx]^2 = [\int_a^b f(x)dx]^2 .$$

而且上式当且仅当在 $\sqrt{g(x)} = c\frac{f(x)}{\sqrt{g(x)}}$ 时（即 $g(x) = cf(x)$ 时）达到极小值

$[\int_a^b f(x)dx]^2$．这说明 $Var(\hat{I}^{(g)}) = \frac{1}{N}[\int_a^b [\frac{f(x)}{g(x)}]^2 g(x)dx - I^2]$ 的最小值在 $g_0(x) \stackrel{\Delta}{=} cf(x)$

时取到．又因为 $g_0(x)$ 为密度，故 $c = \frac{1}{\int_a^b f(x)dx}$． 对应的 $Var(\hat{I}^{(g_0)}) = 0$．此时

$\hat{I}^{(g_0)} = \frac{1}{c} = \int_a^b f(x)dx$ 是无估计误差的精确值．

综上讨论可知，要使方差达到最小，就应该用 $g_0(x) \stackrel{\Delta}{=} cf(x)$ 作为参考密度， 这时用 Von Neumann 取舍方法必须知道参考密度 $g_0(x)$，而计算 $g_0(x)$ 又必须知道常数 c，后者却是于积分的倒数，这样的耦合性的存在使这种方法不能真正用于计算积分． 但是由此我们至少得到了以下的认识，即只要密度 $g(x)$ 的形状与被积分的函数近似，用 $\hat{I}^{(g)}$ 作为 $\int_a^b f(x)dx$ 的估计，就会降低方差． 这个想法还可以推广到 $f(x)$ 未必为非负函数的情形． 这就是下面的概念．

定义８．１ 分布密度为 $g(x) = \frac{|f(x)|}{\int_a^b |f(x)|dx}$ 的 $g-$采样，称为（关于 $f(x)$ 的）重要度采样（Importance Sampling)．

重要度采样不能直接通过取舍原则实现．近似地实现重要度采样可以采用下一节中的 Markov 链 Monte Carlo 方法，这个方法同时给出连积分 $\int_a^b |f(x)|dx$ 的近似算法．

在实践中人们也往往按照重要度采样的思路，灵活地寻找常用的已知类型的密度 g，使它在峰值附近与 $|f(x)|$ 较接近，以便达到降低估计的方差的目的．

修正的重要度采样 (Modified Importance Sampling)

对于 $g-$采样．假定有非负函数 $h(x)$，满足

$$a = \int_a^b h(x)g(x)dx > 0 ,$$

而且a 已知．那么我们可以采用$h(x)$ 作为修正乘积因子．显见对于 $g-$随机数V 有

$$I = \int_a^b f(x)dx = \frac{aE[\frac{f(V)}{g(V)}]}{Eh(V)} . \quad (8.\ 6)$$

如果我们先放弃对于估计的无偏性要求，而只要求估计的相合性，则对于 N 个独立的 $g-$随机数$V_1,\cdots,V_N$，我们可以通过比值，构造 $I = \int_a^b f(x)dx$ 的如下的估计量

$$\check{I} \stackrel{\Delta}{=} a \cdot \frac{\frac{f(V_1)}{g(V_1)} + \cdots + \frac{f(V_N)}{g(V_N)}}{h(V_1) + \cdots + h(V_N)} . \quad (8.\ 7)$$

显见它是$I = \int_a^b f(x)dx$ 的相合估计．在某些假定下，它还是渐进无偏的，即

$$\lim_{N\to\infty} E(\check{I}) = I .$$

而且$\check{I}$ 保留了重要度采样的特性，即当$h(x) = c\frac{f(x)}{g(x)}$ 时，$\check{I}$ 就是$I = \int_a^b f(x)dx$．于是，只要当$h(x)$ 与$\frac{f(x)}{g(x)}$近似，就会降低方差．注意对于给定的分布密度$g(x)$，$h(x)$ 应选取得尽量与$\frac{f(x)}{g(x)}$近似．这个修正乘积因子$h(x)$ 是用来再一次降低由于密度$g(x)$ 与被积函数 $f(x)$ 的倍数不够像所带来的失误而设置的．而当$h(x) \equiv 1$时，就退化为是 $g-$采样，这相当于对 $g-$采样不再作修正．

2．Markov 链 Monte Carlo (MCMC)

用 Markov 链的样本，来对不变分布，Gibbs 分布，Gibbs 场，高维分布，或样本空间非常大的离散分布等作采样，并用以作随机模拟的方法，统称为 Markov Chain Monte Carlo (MCMC)方法．这是动态的 Monte Carlo 方法． 由于这种方法的问世，使随机模拟在很多领域的计算中，相对于决定性算法，显示出它的巨大的优越性．而有时随机模拟与决定性算法的结合使用，会显出更多的长处．

MCMC 至少可以用在以下几个层面：

1．用于生成较复杂的随机数：

* 实现对高维分布（或高维格点分布）p 取样，得到p 随机数．

* 是实现重要度采样的一种方法．对$|f(x)|$的重要性采样，就是取得

p（$=p(x)\stackrel{\Delta}{=}\frac{|f(x)|}{\int|f(y)|dy}$）随机数．

2． 实现高维积分（或项数极多的求和）的数值计算（典例是 Gibbs 分布的各种泛函的平均值的计算）．对于 $f(x)\geq 0$，作以$p=p(x)\stackrel{\Delta}{=}\frac{f(x)}{\int f(y)dy}$为极限分布的 Markov 链$X_n$，利用遍历定理可以由这个 Markov 链的一条轨道，得到分布密度$p(x)$的估计，记为$\hat{p}(x)$．再用$\frac{f(x)}{\hat{p}(x)}$作为积分$I=\int f(x)dx$的估计．

3． 用模拟方法估计最可几轨道．例如，如果模拟了 100 条轨道，那么就能以大概率推断，最可几轨道就在这些轨道的邻近． 当统计量的分布未知时，可以用模拟方法从频率估计置信限．

4．用被估参数的 Bayes 分布（参见第 9 章）的取样，来估计参数．

5．求复杂样本空间上函数的极值(模拟退火)．

2．1 Gibbs 采样法（Gibbs sampler）

1．用 MCMC 方法得到 Gibbs 分布的样本与估计 Gibbs 分布

在第 6 章中，我们考虑了在d维的$N-$格点集上的 Ising 模型的 Gibbs 分布 $p_x=\frac{e^{-bH(x)}}{\sum_{h\in S}e^{-bH(h)}}$，由于所涉及的状态空间（全体组态的集合）$S=\{-1,1\}^{\{1,\cdots,N\}^d}$ 非常大（例如，把一幅 256×256 个采样点的黑白图像看成一个组态，则$d=2,N=256$，S 中有 $2^{256\times 256}(=2^{2^{16}}>2^{60000}>10^{18000})$个元素)，这就使得分母中的求和无法实际完成．而 MCMC 方法就是以通过构造一个以这个 Gibbs 分布 $p_x=\frac{e^{-bH(x)}}{\sum_{h\in S}e^{-bH(h)}}$为不变分布的离散时间的 Markov 链 X_n(它就是 Glauber 动力学中的连续时间的 Markov 链的离散时间采样)，作为模拟计算的基点的． 构造的 Markov 链必须易于计算，所以我们要求它的概率转移速率只容许在组态的一个格点上变动．这样的变动方式，称为 Gibbs 方式，这种抽样方法称为 Gibbs 采样法，或者 Gibbs 样本生成法．这个 Markov 链的不变分布正是此 Gibbs 分布p，我们还要求此 Markov 链的转移矩阵满足 $\mathbf{P}^n\xrightarrow{n\to\infty}\mathbf{1}^T$p．这就是说，要求 Gibbs 分布是 Glauber 动力学的极限分布．

于是，当n大时，X_n的一个样本可以近似地认为取自 Gibbs 分布的一个样本，即按此

Markov 链沿任意一条轨道充分发展，就得到 Gibbs 分布的近似取样．

再则，Gibbs 分布的归一化常数(称为配分常数)$\sum_{h\in S} e^{-bH(x)}$，是一个巨大的求和，即一个"离散的"积分． 用随机模拟法计算这个"离散的"积分的最佳随机数正服从 Gibbs 分布(即重要度采样)． 对于 Gibbs 分布的取样，用通常的取舍原则常常并不可行．例如，分别取 $C=1$，$h(x)=e^{-bH(x)}$，而参考密度 $p_0(x)$ 为组态空间上的均匀分布，这时 $e^{-bH(x)}$ 的值常常小得超出计算精度，而求和变量 x 的范围是庞大的组态空间，这就导致求和无法实际计算．所以需要用 Markov 链 Monte Carlo 方法．用 MCMC 方法生成了以 Gibbs 分布为极限分布的 Markov 链 X_n 以后，由遍历定理用 Markov 链的一条轨道，可给出极限分布（Gibbs 分布）的估计：对于充分大的 N，可令

$$\hat{p}_x = \frac{1}{N}(I_{\{x\}}(X_{N+1}) + \cdots I_{\{x\}}(X_{2N}))\text{，} \qquad (8.8)$$

再用 $e^{-bH(x)}$ 除以 Gibbs 分布在 x 处的估计值 $\frac{e^{-bH(x)}}{\hat{p}_x}$，作为配分函数的估计．在理论上这个估计应该与 x 的取法无关．但是， 在实际计算中对多个不同的组态 x_i 分别估计此和数后，再作平均常常能降低方差．

在第 6 章中，我们曾给出了用 Glauber 动力学构造的两个不同的连续时间的 Markov 链(对应于两个不同的转移概率速率矩阵 **Q**)，它们都以 Gibbs 分布 p 为极限分布，而且都是可逆的．

较为深入的理论研究表明，使用不可逆的，且以 p 为不变分布的 Markov 链作 Markov Monte Carlo，会加快这个极限的收敛速度．然而，在另一方面这种做法又会增加计算的复杂程度．再则，为减少估计的方差而作的努力也常会增加计算时间．这就是说，在计算中，我们会面临难以两全的抉择．在实际中如何采取折衷，既要看问题的性质，又要参考实践的经验，没有统一的原则．

用以完成 MCMC 采样操作的 Markov 链，可构造如下：

在第 6 章中，对于 d 维有限格点上，由具有两个自旋的组态空间上的能量函数

$$H(h) = -\frac{1}{2}\sum_{x,y\text{相邻}} h(x)h(y) - h\sum_{x} h(x)\text{，} \qquad (8.9)$$

可构造如下的转移概率速率

$$q_{xh} = \begin{cases} C(x,x), & (h = x^{x}) \\ 0 & (h \neq x\text{的其它情形}) \end{cases} \qquad (8.10)$$

($C(x,x)$ 的两种取法各为：

$$C(x,x) = e^{-b(H(x\ )-H(x))} \qquad (8.11)$$

或

$$C(x,x)=\frac{1}{1+e^{-b\,(H(x^x)-H(x))}}\text{)}\ .\tag{8.11'}$$

它们决定的Glauber 动力学，分别对应的两个不同的连续时间的Markov链，它们都以$H(h)$为能量函数的 Gibbs 分布p 为可逆不变分布，而且p 还是转移矩阵的极限分布：$\mathbf{P}(t)\xrightarrow{t\to\infty}=\mathbf{1}^T p$ ．如果考虑间隔为Δt 的采样时间，其中Δt 充分小，我们还可以进行如下的近似计算．假定初始组态为V ．在时刻Δt 它以概率$C(x,V)\Delta t$ 转移到组态V^x ，而以概率 $1-\Delta t\sum_{x\in\{1,\cdots,N\}^d}C(x,V)$ 停留在原来的组态，这就近似地得到了 Markov 链在Δt 时刻所处的组态V_1．再以类似的方式得到 Markov 链在$2\Delta t$ 时刻所处的组态V_2 ，依次下去．当采样数n 充分大时，组态h 在这段$\{V_1,\cdots,V_n\}$中出现的频率，就用作p_h 的估计$\hat{p}_h$ ．

Gibbs 分布的一些统计平均量的模拟近似

对于上面定义的，以 Gibbs 分布为极限的 Markov 链，我们用它的一段样本$V_1,\cdots,V_N$，可以给出如下的 Gibbs 统计平均量

$$F=\frac{\sum_{x\in S}f(x)e^{-bH(x)}}{\sum_{x\in S}e^{-bH(x)}}\tag{8.12}$$

的模拟近似

$$\hat{F}=\frac{\sum_{i=1}^{n}f(V_i)}{n},\tag{8.13}$$

它是F 的无偏估计．不难证明：存在数$V(f,P,p)<\infty$，使$\hat{F}$ 的方差满足

$$Var(\hat{F})=\frac{V(f,P,p)}{n}+o(\frac{1}{n})\le\frac{13\|f\|^2}{(1-C(P))n},\tag{8.14}$$

其中 $\|f\|=\sum_x f(x)$，而$C(P)$ 为此 Markov 链的转移矩阵$\mathbf{P}$ 的 Dobrushin 收缩数．

Gibbs 采样法还可以有效地用于 Bayes 方法中的后验密度的采样的模拟计算(见第９章)．

2．2 Metropolis 采样法 (Metropolis sampler)

1．Gibbs 分布的采样的 Markov 链 Metropolis 方法

为了突出主要思想，下面我们把组态空间(状态空间)$S=\{-1,1\}^{\{1,\cdots,N\}^d}$ 简单地记为$\{1,\cdots,K\}$．对于组态空间上给定的分布p ， Metropolis 构造了以

$$p_{ij} = \tilde{p}_{ij} \min(1, \frac{p_j}{p_i}), \qquad (j \neq i) \qquad (8.15)$$

（ $p_{ii} = 1 - \sum_{j \neq i} p_{ij}$ ）为转移的时齐 Markov 链，其中$\tilde{\mathbf{P}} = (\tilde{p}_{ij})$是一个对称的互通转移矩阵，称为预选矩阵，或访问方案，使用它是为了减少或简化状态间的连接，以加快 Markov 链的分布向不变分布收敛的速度. 显见p 是它的可逆分布（注意在 Gibbs 分布情形，状态i, j体现为组态x, h， 于是在计算（８．１５）式的转移概率时，就只需算比值

$$\frac{p_x}{p_h} = e^{-b(H(x)-H(h))} \qquad (x \neq h)$$
，而并不需要计算配分函数． Glauber 动力学的构架，也正是用了这一点）． 由这个有限状态 Markov 链的互通性，我们有

$$\mathbf{P}^n \xrightarrow{n \to \infty} = \mathbf{1}^T p \ .$$

因此，在时间发展充分长以后，我们可以用 Metropolis 的 Markov 链所处的状态，作为按分布p 取的样本．也就是说， 与 Gibbs 采样法一样，Metropolis 方法也给出了在计算机上模拟p - 随机数的一个算法．Metropolis 提出的这种采样法，称为 Metropolis 采样法．它与 Gibbs 采样法的不同处在于，对于 Metropolis 采样法而言，任意两个组态x, h， 只要预选概率$\tilde{p}_{xh} > 0$就可以转移．

Metropolis 采样在时刻n的更新$x^{(n)} \to x^{(n+1)}$可具体采取如下的操作:

(1) 设当前为时刻n，取的状态为$x^{(n)} = i$．对它作随机扰动，即取一个分布为$(\tilde{p}_{i1}, \cdots, \tilde{p}_{iK})$的随机数，设为$j$；

(2) 若$\frac{p_j}{p_i} \geq 1$，则将状态更新为$x^{(n+1)} = j$；否则进行（３）；

(3) 独立地取一个$U[0,1]$随机数U，如果$U \leq \frac{p_j}{p_i}$，则将状态更新为$x^{(n+1)} = j$；否则状态不更新，即令$x^{(n+1)} = i$．

（请读者证明，如此由i到j的转移的可能性恰是（８．１５）式规定的转移概率）．

$\tilde{\mathbf{P}}$的对称性并非必要. 理论分析指出，经过适当的选取（研究矩阵$\tilde{\mathbf{P}}$的第二个特征值），使用非对称的 $\tilde{\mathbf{P}}$ 可能加快收敛速度. 对于非对称的预选矩阵$\tilde{\mathbf{P}}$，Metropolis 采样法所构造的 Markov 链的转移应取下式：

$$p_{ij} = \tilde{p}_{ij} \min(1, \frac{p_j \tilde{p}_{ji}}{p_i \tilde{p}_{ij}}), \quad (j \neq i). \qquad (8.16)$$

此时所构造的 Markov 链仍然以p 为可逆分布（请读者验证，并给出这种 Metropolis 采样在时刻n 的更新 $x^{(n)} \to x^{(n+1)}$ 可具体采取如下的操作步骤）. .

具有非对称的预选矩阵的 Metropolis 采样法的一个特殊情形就是独立的 Metropolis 采样，即预选转移矩阵与出发点无关的 Metropolis 采样．此时预选矩阵各行是一样的，因此可以写为

$$\tilde{\mathbf{P}} = \mathbf{1}^T \mathrm{r}.$$

记 $w_j = \frac{p_j}{r_j}$，结合拒收原则，对独立的 Metropolis 采样实际操作为

（1） 按分布 r 采样，得到 j ；

（2） 若 $\frac{w_j}{w_{x^{(n)}}} \geq 1$，则取$x^{(n+1)} = j$；否则进行(3);

（3） 取$U[0,1]$ 随机数U ， 修改$x^{(n)}$ 为 $x^{(n+1)}$ ：

$$x^{(n+1)} = \begin{cases} j & (U \leq \min(1, \frac{w_j}{w_{x^{(n)}}}) \\ x^{(n)} & (\text{其他情形}) \end{cases}。$$

注１ 一般地 Markov 链的两个状态之间都可能转移， 因此 Markov 链相当于连续取值情形中的"高维情形"．预选矩阵的设计是尽可能地减少，或者简化状态间的连接，以减少计算量，有时它相当于局部化．

注２ 在独立的 Metropolis 采样法，且p 是一维密度的情形，我们还可以把 Metropolis 算法视为 Von Neuman 拒收原则在某种意义下的推广．此时拒收原则中的$p_0(x)$就相当于预选矩阵． 为了方便做比较，我们把p 改记为 $p(x)$．此时独立的 Metropolis 算法可以设计如下：取一个与 分布密度 $p(x)$ 取值范围差不多的随机数的独立样本 $h_0, h_1, \cdots$，记其分布密度为 $p_0(x)$， 再取与它独立的$U[0,1]$ 独立随机数列 $U_1, U_2, \cdots$．令$x_0 = h_0$，然后归纳地定义

$$x_{n+1} = \begin{cases} h_{n+1} & (U_n \leq \min(\frac{p(h_{n+1})}{p(x_n)} \frac{p_0(x_n)}{p_0(h_{n+1})}, 1) \\ x_n & (\text{其它情形}) \end{cases}.$$

于是当n 大时，x_n 就近似地是p 随机数．(这里 $p_0(x)$起了 p_{ij} 的作用)．把它与拒收原则相

比较可以看出： Metropolis 算法不需要假定 $\frac{p(x)}{p_0(x)}$ 的有界性，为此付出的代价是：不能从是否有 $U \le \frac{p(h)}{Cp_0(h)}$ 来决定是否将 x 取为 h，代替一个更新的随机数 h 以一个随机数的近似序列 h_n， 并且从是否有

$$U_n \le \text{ 序列} \frac{p(h_{n+1})}{p_0(h_{n+1})} \text{与其前面的值} \frac{p(x_n)}{p_0(x_n)} \text{的比}$$

来决定是否把 x_n ”更新为” h_{n+1}． 所以可以说，独立的 **Metropolis** 算法主要用在找不到 $p_0(x)$ 使 $\frac{p(x)}{p_0(x)}$ 有界的情形，特别是在 $p(x)$ 只能得到测量的数据值，而没有显式表示的情形． 然而，在实际计算中一般地只在 " 高维情形 " 才有必要使用 Metropolis 算法．

注３ 比较这 Gibbs 采样法与 Metropolis 采样法的收敛速度，是一个重要的理论问题．

注４ 使用非对称的 $\tilde{\mathbf{P}}$ 的另一种 Markov 转移的取法，是如下的 Hastings 推广了的 Metropolis 采样

$$p_{ij} = \tilde{p}_{ij} \frac{s_{ij}}{1 + \frac{p_j \tilde{p}_{ji}}{p_i \tilde{p}_{ij}}}, \quad (j \ne i), \tag{8. 17}$$

其中 $\tilde{\mathbf{P}}$ 为非对称的转移矩阵，满足: $\tilde{p}_{ij}$ 与 $\tilde{p}_{ji}$ 要么同时为零，要么同时为正，而 (s_{ij}) 是一个非负的对称矩阵，满足

$$\frac{s_{ij}}{1 + \frac{p_j \tilde{p}_{ji}}{p_i \tilde{p}_{ij}}} \le 1, \quad (j \ne i). \tag{8. 18}$$

以上几种采样法都需要知道 p 的解析表达式．在并不知道它的表达式时，下面的方法对于采集以它为分布的样本，能起很好的作用．

2．3 通过条件分布对分布 p 作随机采样的 Gibbs 方法

Gibbs 采样法中 Markov 链的状态的更新按 Gibbs 方式进行，即更新时只变动组态的一个分量，它实行起来较为简单，所以这种思想可以用来得到高维总体的样本．记高维总体的分布为 p （未必是 Gibbs 分布）．Gibbs 方式更新的思想，在用于甚高维总体或复杂总体的取样时，主要是通过 p 的条件分布族，来构造一个不可约正常返的马氏链 X_n，使它以 p 为不变分布．由于 n 充分大时，X_n 的分布渐近于 p ， X_n 就可以近似地作为分布 p 的总体

的样本． 下面我们以可数状态空间的分布p 为例来进行讨论，其处理问题的思想，也适用于生成有密度的总体的样本．

1 由条件分布族得到以p 为不变分布的 Markov 转移的思想

Gibbs 样本生成法的要义，在于把高维总体的取样，化为一系列一维分布的取样，而后者是容易做到的．甚至可以利用 SAS 等软件得到．但是，在具体运作时，对于较大的m及一个给定的 m 维分布p ，事情就远非如此简单．设$p=p(x_1,\cdots,x_m)$．记$x=(x_1,\cdots,x_m)$，$y=(y_1,\cdots,y_m)$．再记在第 1 至 k-1 个分量固定为$y_1,\cdots,y_{k-1}$，同时将第 $k+1$ 至第m 个分量固定为 $x_{k+1},\cdots,x_m$的条件下，第 k 个分量在y_k的条件概率为

$$p(y_k \mid y_1,\cdots,y_{k-1},x_{k+1},\cdots,x_m).$$

我们定义如下的转移概率

$$p_{xy}\overset{\Delta}{=}p(x,y)=\prod_{k=1}^{m} p(y_k \mid y_1,\cdots,y_{k-1},x_{k+1},\cdots,x_m). \tag{8.19}$$

容易验证(p_{xy})是一个转移矩阵．我们证明p 是它的不变分布．事实上

$$\sum_x p(x)p(x,y)=\sum_x p(x)\prod_{k=1}^{m} p(y_k \mid y_1,\cdots,y_{k-1},x_{k+1},\cdots,x_m)$$

$$=\sum_{x_2,\cdots,x_m}[\sum_{x_1} p(x_1,\cdots,x_m)]\frac{p(y_1,x_2,\cdots,x_m)}{\sum_{x_1}p(x_1,\cdots,x_m)}\prod_{k=2}^{m} p(y_k \mid y_1,\cdots,y_{k-1},x_{k+1},\cdots,x_m)$$

$$=\sum_{x_2,\cdots,x_m}p(y_1,x_2,\cdots,x_m)\prod_{k=2}^{m} p(y_k \mid y_1,\cdots,y_{k-1},x_{k+1},\cdots,x_m)$$

$$=\sum_{x_3,\cdots,x_m}[\sum_{x_2} p(y_1,x_2,\cdots,x_m)]\frac{p(y_1,y_2,x_3,\cdots,x_m)}{\sum_{x_2}p(y_1,x_2,\cdots,x_m)}\prod_{k=3}^{m} p(y_k \mid y_1,\cdots,y_{k-1},x_{k+1},\cdots,x_m)$$

$$=\sum_{x_3,\cdots,x_m}p(y_1,y_2,x_3,\cdots,x_m)\prod_{k=3}^{m} p(y_k \mid y_1,\cdots,y_{k-1},x_{k+1},\cdots,x_m)$$

$$=\cdots=p(y_1,\cdots,y_m)=p(y).$$

［注］ 事实上，我们并不需要知道$p(x)$的表达式，而只需知道在固定其它一切分量的条件下，余下的某一个分量的条件分布，就可以得到以$p(x,y)$为转移的时齐马氏链的样本．

2．构造 Markov 的方法

从已得到的 Markov 链在时刻n的样本X_n，去求转移到时刻$n+1$的样本X_{n+1}时, 转移$p(x,y)$是一维的条件分布的乘积． 而这些一维条件分布的样本是较为容易生成的, 即可

按以下程序逐个地得到 X_{n+1} 的样本的各个分量 $(y_1,\cdots,y_m)$：

（1） 先得到服从分布 $\{p(y_1\mid x_2,\cdots,x_m), y_1\in S\overset{\Delta}{=}\{1,\cdots,N\}\}$ 的随机变量（ 记为 $X_{n+1,1}$） 的一个样本 y_1 （在用 Von Neumann 取舍原则取样时，对于 $p(y_1\mid x_2,\cdots,x_m)=\dfrac{p(y_1,x_2,\cdots,x_m)}{p(+\infty,x_2,\cdots,x_m)}$，只需假定对固定的 $(x_2,\cdots,x_m)$，存在 $p_0(y_1)$ 及常数 C，满足 $p(y_1,x_2,\cdots,x_m)\le Cp_0(y_1)$，并对 $U[0,1]$ 均匀随机数 U，看 $\dfrac{p(U,x_2,\cdots,x_m)}{Cp_0(U)}$ 是否不小于 U，来决定取舍样本 U）.

（2）用同样的方法，再得到服从分布 $\{p(y_2\mid y_1,x_3,\cdots,x_m), y_2\in S\overset{\Delta}{=}\{1,\cdots,N\}\}$ 的随机变量（ 记为 $X_{n+1,2}$） 的一个样本 y_2.

（3）依此下去，得到服从分布 $\{p(y_k\mid y_1,\cdots,y_{k-1},x_{k+1},\cdots,x_m), y_k\in S\}$ 的随机变量（ 记为 $X_{n+1,k}$） 的一个样本 y_k，$(k\le m-1)$.

（4）最后得到服从分布 $\{p(y_m\mid y_1,\cdots,y_m), y_m\in S\}$ 的随机变量（ 记为 $X_{n+1,m}$） 的一个样本 y_m. 于是 $y=(y_1,\cdots,y_m)$ 就是 的一个样本.

现在我们任取 $X_0=y(0)$，再按上面方法得到 X_1 的一个样本 $y(1)$. 对 n 归纳地我们可用上面方法分别得到 $X_2,\cdots,X_n$ 的样本 $y(2),\cdots,y(n)$. 当 n 充分大时，由于 Markov 链 X_n 的分布近似于 $p(*)$. 我们就可以认为 $y(n)$ 近似地是分布 $p(*)$ 的一个样本.

[注] 上面的 Gibbs 采样也可以看成为如下的算法

（1）按循环次序更新状态的各个分量，在更新了第 d 个分量后， 就接着更新第一个分量；

（2）按条件分布 $p(\bullet\mid x_1^{(n)},\cdots,x_{j-1}^{(n)},x_{j+1}^{(n)},\cdots,x_d^{(n)})$ 选取 $x_j^{(n+1)}$，再把状态 $(x_1^{(n)},\cdots,x_d^{(n)})$ 更新为 $(x_1^{(n+1)},\cdots x_j^{(n+1)},\cdots,x_d^{(n+1)})\overset{\Delta}{=}(x_1^{(n)},\cdots,x_{j-1}^{(n)},x_j^{(n+1)},x_{j+1}^{(n)},\cdots,x_d^{(n)})$。

[注] 回顾 Gibbs 采样，还可以从理论角度得到以下认识：如果状态空间是实数，我们把 Gibbs 采样中第 k 个分量的条件更新看成 x_k 到 $y_k=x_k+x_k$ 的随机(加法)变换：$y_k=T_{x_k}x_k$，那么 Gibbs 采样的更新就是说，. 只要 x_k 服从条件分布 $p(y_k\mid y_1,\cdots,y_{k-1},x_{k+1},\cdots,x_n)$，则目标函数(即不变密度 $p(x)$)在这类随机变换下是不变的. 这种看法有助于进一步设计有效

的 MCMC 采样方法.

3　Gibbs 采样法与 Metropolis 采样法的变种应用

第１种变种（ 随机变种）

这时，状态 $(x_1^{(n)},\cdots,x_d^{(n)})$ 的更新，不按指标次序轮番更替，代之以

（1）　每次按某个"整值"随机变量 h 的采样值 j，来随机选取指标；

（2）　再按条件分布 $p(*\mid x_1^{(n)},\cdots,x_{j-1}^{(n)},x_{j+1}^{(n)},\cdots,x_d^{(n)})$ 选取 $x_j^{(n+1)}$；

（3）　然后把 $(x_1^{(n)},\cdots,x_d^{(n)})$ 更新为：

$$(x_1^{(n+1)},\cdots x_j^{(n+1)},\cdots,x_d^{(n+1)})\overset{\Delta}{=}(x_1^{(n)},\cdots,x_{j-1}^{(n)},x_j^{(n+1)},x_{j+1}^{(n)},\cdots,x_d^{(n)})。$$

第２种变种（分裂变种）

设状态变量 $\vec{x}$ 可以分为两部分 $\vec{x}=(\vec{x}_1,\vec{x}_2)$，其中 $\vec{x}_2$ 是丢失数据的部分. 分别记这两部分随机变量相互间的条件分布为 $p_{2|1}(\vec{x}_2\mid\vec{x}_1)$，$p_{1|2}(\vec{x}_1\mid\vec{x}_2)$. 如果我们可能在 $\vec{x}_1$ 已知时对 $p_{2|1}(\vec{x}_2\mid\vec{x}_1)$ 进行采样，也可能在 $\vec{x}_2$ 已知时对 $p_{1|2}(\vec{x}_1\mid\vec{x}_2)$ 进行采样. 那么，我们可以按下面步骤进行：

(1)　任意设置一个初值，例如 $\vec{x}_1^{(0)}$；

(2)　由条件分布 $p_{1|2}(*\mid\vec{x}_2^{(n)})$ 中抽取 $\vec{x}_1^{(n+1)}$；

(3)　再由条件分布 $p_{2|1}(*\mid\vec{x}_1^{(n+1)})$ 中抽取 $\vec{x}_2^{(n+1)}$，这样依次交替进行.

只要 $p_{2|1}(\vec{x}_2\mid\vec{x}_1)$，$p_{1|2}(\vec{x}_1\mid\vec{x}_2)$ 恒正，那么由上面的 Gibbs 采样的一般理论可知，$(\vec{x}_1^{(n)},\vec{x}_2^{(n)})$ 就是一个 Markov 链的采样值，而且 $(\vec{x}_1,\vec{x}_2)$ 的联合分布是这个链的不变分布. 于是当 n 充分大的时候， 可以认为 $(\vec{x}_1^{(n)},\vec{x}_2^{(n)})$ 是近似地采自 $(\vec{x}_1,\vec{x}_2)$ 的联合分布的一个样值。

第３种变种（Metropolis 采样型变种）

（1）每次按某个随机变量 h 的采样值 j，来随机选取指标；

（2）再按条件分布 $p(\bullet\mid x_1^{(n)},\cdots,x_{j-1}^{(n)},x_j\neq x_j^{(n)},x_{j+1}^{(n)},\cdots,x_d^{(n)})$ 选取一个样本 z；

（3）取一个与之独立的 $U[0,1]$ 样本 U， 令

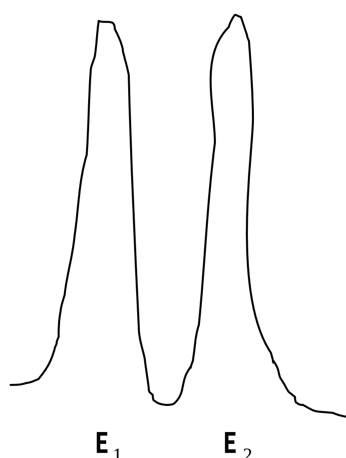
$$x_j^{(n+1)}=\begin{cases} z & \left(U\le\min\left(1,\dfrac{p(x_j^{(n)}\mid x_1^{(n)},\cdots,x_{j-1}^{(n)},x_{j+1}^{(n)},\cdots,x_d^{(n)})}{p(z\mid x_1^{(n)},\cdots,x_{j-1}^{(n)},x_{j+1}^{(n)},\cdots,x_d^{(n)})}\right)\right. \\ x_j^{(n)} & (\text{其他情形}) \end{cases}；$$

（4）然后把 $(x_1^{(n)},\cdots,x_d^{(n)})$ 更新为

$$(x_1^{(n+1)},\cdots x_j^{(n+1)},\cdots,x_d^{(n+1)})\overset{\Delta}{=}(x_1^{(n)},\cdots,x_{j-1}^{(n)},x_j^{(n+1)},x_{j+1}^{(n)},\cdots,x_d^{(n)})。$$

4.利用可逆链防止多峰情形的偏取

在 Gibbs 样本生成过程中，只有当马氏链{ X_n }发展的"时间" n 充分大时，它的分布才近似于 p ．在实践中怎样才能判断 n 是否已经足够大了呢？ 通常的判断标准一般是看在计算中的 X_n 的分布是否比较稳定了．但是，在事实上这种判断方法不尽正确．下面的示意图给出了 $p=p(x)$ 具有两峰的情形．从不同初值出发的轨道，可能分别稳定在 E_1 与 E_2 两个区域，而事实上 X_n 还远未达到"近似于"不变分布的要求．



E_1 E_2

也就是说，按上一段的方法得到的 Markov 链，从不同的初值出发，在计算达到看起来稳定时，它们还远不能近似π(x)的样本．原因在于这时的平稳分布π(x)有两个峰，且在两峰中间的转移概率非常小，以使从 E_1 中的点出发能到达 E_2 和从 E_2 中的点出发能到达 E_1 的概率都极为微小，从而在一条轨道上要从 E_1 到达 E_2 就要化费十分长的时间．有时甚至在现实生活中可能不会在有意义的时间内达到．发现与使用Gibbs样本生成法的先驱们往往希望通过选取多个初始值，得到多条轨道以解决这一问题．但是这并不能完全解决问题．正如在当前的例子中的两条轨道各自只局限在不同的区域 E_1 和 E_2 中，它们的分布，根本反映不出分别限制在 E_1 和 E_2 上的 $p=p(x)$ 如何拼接的， 即以多大的比例 $C_1:C_2$ ，使

$$(C_1+C_2)\,p\approx C_1\,p\,|_{E_1}+C_2\,p\,|_{E_2}\,.$$

只有找到了这个比例 $C_1:C_2$ ，才能从分别限制在 E_1 和 E_2 上的 $p\,|_{E_1}$ 和 $p\,|_{E_2}$ ，拼接出整体的π(x)．

为了找出比例 $C_1:C_2$ ，我们建议利用可逆 Markov 链来生成 Gibbs 样本，并从可逆 Markov 链的特性得到不变分布在它的各个峰区上的比例．下面我们以双峰时两条轨道的情

形，来阐述确定与估计这个比例的方法．

设分别稳定在 E_1 和 E_2 上两条轨道的样本分布对应于分布 $p_1(x)$ 与 $p_2(x)$．我们近似地把 $p(x)$ 在 E_1 和 E_2 外为 0．于是由可逆 Markov 链的性质可知，它们应该分别与{π(x):x∈E_1}及{π(x):x∈E_2}成比例，即

$$\frac{p_i(x)}{p_i(y)}=\frac{p(x)}{p(y)},(\forall x,y\in E_i)\text{．}$$

于是

$$\frac{\hat{p}_i(x)}{p(x)}\approx\frac{p_i(x)}{p(x)}\stackrel{\Delta}{=}C_i\text{．}$$

因此

$$C_i\approx\frac{1}{p(E_i)}\text{．}$$

具体确定 C_i 可如下进行：对 $x_i\in E_i$，利用可逆性我们有

$$\frac{C_1}{C_2}=\frac{p_1(x_1)p(x_2)}{p_2(x_2)p(x_1)}=\frac{p_1(x_1)p(x_1,x_2)}{p_2(x_2)p(x_2,x_1)}\text{．}$$

于是对 $x,y\in E_i$，有

$$1=\sum_{x\in E_i}p_i(x)=p_i(y)\sum_{x\in E_i}\frac{p(y,x)}{p(x,y)}=p_i(y)E_y[\frac{1}{p(X_1,y)}]\text{．}$$

对于 X_1 的 N 次取样 $X_1^{(1)},\cdots,X_1^{(N)}$，我们可以得到 $p_i(y)=\left(E_y[\frac{1}{p(X_1,y)}]\right)^{-1}$ 的估计

$$\hat{p}_i(y)=\left(\frac{1}{N}\sum_{k=1}^{N}\frac{1}{p(X_1^{(k)},y)}\right)^{-1}\text{．}$$

这里的近似只需对从 y 出发的 Markov 链，作独立的 N 次一步转移的模拟．它比 Markov 链的分布近似于不变分布要求的步数要少得多．

余下的问题是从不变分布π(x)具体构造可逆 Markov 链．这只需把前面构造的马氏转移 $p(x,y)$ 修正为

$$\hat{p}(x,y)=\frac{1}{2}\left(p(x,y)+\frac{p(y)p(y,x)}{p(x)}\right).\qquad(8.20)$$

容易检查它是以 $p(x)$ 为可逆分布的转移阵．这恰是我们所希望要的．

为了便于具体操作，我们还希望 $\hat{p}(x,y)$ 也按 Gibbs 方式转移，即也能归结于计算每次只变动一个分量的 m 个“条件转移概率”的乘积．为此我们注意：

$$\frac{p(y)p(y,x)}{p(x)}$$

$$=\frac{p(y_1,\cdots,y_m)p(x_1\mid y_2,\cdots,y_m)p(x_2\mid x_1,y_3,\cdots,y_m)\cdots p(x_m\mid x_1,\cdots,x_{m-1})}{p(x_1,\cdots,x_m)}$$

$$=\frac{p(y_1,\cdots,y_m)}{p(x_1,\cdots,x_m)}\frac{p(x_1,y_2,\cdots,y_m)}{\int p(y,y_2,\cdots,y_m)dy}\cdots\frac{p(x_1,\cdots,x_{m-1},y_m)}{\int p(x_1,\cdots,x_{m-2},y,y_m)dy}$$

$$\times\frac{p(x_1,\cdots,x_m)}{\int p(x_1,\cdots,x_{m-1},y)dy}$$

$$=p(y_1\mid y_2,\cdots.y_m)\cdots p(y_{m-1}\mid x_1,\cdots,x_{m-1},.y_m)\,p(y_m\mid x_1,\cdots.x_{m-1})\text{．}$$

即 $\frac{p(y)p(y,x)}{p(x)}$ 具有类似于 $p(x,y)$ 的形式,不同处只在于所有条件分布中的变量次序倒了一下．而 $\hat{p}(x,y)$ 的含意为：分别以 $\frac{1}{2}$ 的概率取 $p(x,y)$ 与取 $\frac{p(y)p(y,x)}{p(x)}$．所以，从模拟程序的实现来看，这仅仅增加了一倍计算量．

2. 4 MCMC 应用于 Bayes 参数估计

1 关于 Bayes 统计

我们将在第９章中较多地介绍与评述 Bayes 统计．Bayes 统计方法的基本想法是： 把概率函数中的未知参数 q 当作随机变量.

由于在 Bayes 方法中不再区分参数与随机变量,所以未知参数的分布的确认是至关重要的．在抽取样本以前，就只能根据先验知识, 设置未知参数的分布，称为先验分布。在取样以后，根据对于抽取到的样本的概率规律的了解，使用 Beyes 公式，把参数的先验分布改进为后验分布，后验分布也称 Bayes 分布。

设随机变量 x 的分布为 $p(x,q)$（未知参数 q 可以是向量)．把 q 看成随机变量（或随机向量）后，$p(x,q)$ 就是 :在 q 已知的条件下，x 的条件分布， 即 $p_{x|q}(x\mid q)=p(x,q)$． 假定 q 的先验分布的分布密度为 $f(J)$. 于是 (x,q) 的联合分布为 $f(q)\,p(x,q)$．

在 J 固定条件下, 设 $x_1,\cdots,x_n$ 为采自 $p(x,q)$ 的独立随机样本. 把在 $x_1,\cdots,x_n$ 已知的条件下 q 的条件分布，记为 $j(J\mid x_1,\cdots,x_n)$．那么，它等于 q 与 $(x_1,\cdots,x_n)$ 的联合密度

$f(q)\prod_{i=1}^{n} p(x_i,q)$，除以$(x_1,\cdots,x_n)$的边缘密度，即

$$j(J\,|\,x_1,\cdots,x_n) = \frac{f(J)\prod_{i=1}^{n} p(x_i,J)}{\int f(J)\prod_{i=1}^{n} p(x_i,J)dJ}.$$

这就是q的 Bayes 分布．所以至关重要的是如何选取参数的先验分布，这个问题将在第９章中作一些较为浅显的介绍．

2　用 MCMC 方法对 Bayes 分布参数作估计

对于参数q的 Bayes 分布， 用 MCMC 方法做以它为极限分布的 Markov 链J_n．于是,

由于J_n的分布近似于此 Bayes 分布, 我们可以取$\hat{J} = \dfrac{J_{m+1}+\cdots+J_{m+N}}{N}$, $(m \geq n_0$， n_0充分大) 作为参数J估计.

３．用 MCMC 方法得到较为复杂的参数估计的大意

较复杂的参数估计的典型情形是混合总体的参数 Bayes 估计. 设样本$X_1,\cdots,X_n$来自权重参数未知的混合总体，这些未知权重也就作为另加的待估计参数．在对混合分布作参数估计时, 可以引进一些新的随机变量 (在 Bayes 方法中，原则上参数与随机变量不再严格区分, 可以把它们看成“补偿参数”) Z, 然后用 Gibbs 采样法， 交替使用：在ϑ已知的条件下，Z的条件分布$p_{2|1}(z\,|\,\vartheta)$，与在Z已知的条件下，ϑ的条件分布$p_{1|2}(\vartheta\,|\,z)$。由此可得到(ϑ, Z)的 Bayes 分布的采样，其步骤如下：

(1) 任给初值ϑ_0；

(2) 按条件分布$p_{2|1}(z\,|\,\vartheta_n)$采样， 得样本$z_n$；

(3) 按条件分布$p_{1|2}(\vartheta\,|\,z_n)$采样， 得更新样本$\vartheta_{n+1}$。

于是只要$p_{2|1}(z\,|\,\vartheta)$与$p_{1|2}(\vartheta\,|\,z)$恒正， 那么就有：(ϑ_n, z_n)是 Markov 链， 而且以(θ, Z)的联合分布为极限分布。从而当m大时，可以认为(ϑ_m, z_m)近似地是出自(θ, Z)的联合分布的采样，同时可以用取$\hat{J} = \dfrac{J_{m+1}+\cdots+J_{m+N}}{N}$作为$\vartheta$的估计量。

以上的关键在于知道$p_{2|1}(z\,|\,\vartheta)$与$p_{1|2}(\vartheta\,|\,z)$， 而且易于抽取样本，它们还不能以正概率取 0，以保证所作的 Markov 链的转移函数的收敛性． 关于$p_{2|1}(z\,|\,\vartheta)$与$p_{1|2}(\vartheta\,|\,z)$的取法与获得， 在现代一些统计论文中，有一些典型的做法，本书不再叙述．

4 MCMC 方法用于缺失数据情形下的参数估计的大意

如果 $\vec{X}=(\vec{Z},\vec{Y})$ 的分布密度有未知参数 $\vec{\vartheta}$ ： $p(\vec{z},\vec{y},\vec{\vartheta})$ ， 其中 $\vec{Y}$ 是可以通过采样观测到的， 但是, $\vec{Z}$ 却观测不到（或者是由数据的缺失而引起的，它称为缺失数据）. 用 Bayes 统计的思想， 在观测样本 $\vec{Y}$ 条件分布给定的条件下， 记条件分布为

$$p_{2|1}(z\mid\vartheta)=p(\vec{z},\vec{Y},\vec{\vartheta}),\ \ p_{1|2}(\vartheta\mid z).$$

只要它们恒正， 而且有一种方法可以近似地取得服从这两个分布的样本，那么就可以用第３段中的方法得到缺失数据 $\vec{Z}$ 的估计.

3. 模拟退火 (simulated annealing)

３．１ 模拟退火方法的基本想法

一般的优化问题，都可以归结为求某个目标函数 f 在一定范围内的最小值. 但是许多常用的方法，往往都是运用某种局部的比较去决定优化过程的发展，以逐步达到最小值. 例如梯度法，就是每次将目标函数在当时所考察的点与附近的点的取值加以比较，根据梯度方向就是目标函数值下降最快的方向，导出了梯度下降法，即在被考察的点沿函数下降最快的方向移动，希望经过多次这样的移动，最终达到目标函数 f 最小的点. 但是这个算法的实际结果是停止在梯度为零的某个点上. 这就是说，在计算进行到到达目标函数的某个局部极小值点时，在考察点的目标函数值就小于它附近所有点的目标函数值，于是搜索过程就停止了. 我们知道目标函数的任意一个局部极小值点都满足上述条件，然而局部极小值点处的函数值,有可能比函数的实际最小值大得多.于是求得最小值的努力就完全寄托于初值的选取.

为了克服上述算法的这个缺点，Metropolis 等人在 1953 年提出的一种有效地模拟固体达到热平衡发展的算法. 而 Khasi′minski 在 1965 年对决定性的梯度算法，引入小的随机噪声干扰，使之在达到平衡时，以高概率进入目标函数的最小值附近，但是当时没有受到重视. Kirkpatrick 等人于 1983 年对组合优化问题，重新提出模拟退火算法. 其基本想法正是 Khasi′minski 的构想. 即引入人工噪声使得当算法达到局部极小值时，能有一个小概率“逸出”局部极小值的陷井. 这个想法源自冶金退火过程的物理思想，即将金属加温至充分高，再使其慢慢地退火冷却. 在加温时，金属内部粒子随温度增加变为无序，能量增大; 而在退火时，在慢慢冷却的过程中粒子渐趋于有序，在每个温度时都达到平衡态，最后在常温时达到能量最大的基态. 而用退火模拟优化问题时，则将能量模拟为目标函数，温度演化成代表噪声水平的控制参数 b ，这样就得到优化问题的模拟退火算法.

简单而实用的模拟退火模型是有限状态模型，典型情形是组合优化问题. 这时假定自变量只能取自有限个值. 而加入随机噪声的算法，就是把原来的计算过程变成随机过程 — 也就是一个 Markov 链，并使得此 Markov 链是互通的，非周期的，正常返的，且具有如下形式的不变分布:

$$p_i^{(b)} = \frac{1}{\sum_j e^{-b(f(j)-f_*)}} e^{-b(f(i)-f_*)} = \frac{1}{\sum_j e^{-b\cdot f(j)}} e^{-b\cdot f(i)}, \qquad (8.20)$$

其中 f 是我们要优化的目标函数，$f_* = \min_i f(i)$，而 b 是一个控制噪声水平(随机性水平)的参数．这个分布有一个重要的性质：当 $b \to \infty$ 时，其极限分布集中在目标函数 f 取最小的集合上，即在 $\{i : f(i) = f_*\}$ 上（f 取最小值的点可能不唯一）．于是我们可以利用这个 Markov 链的发展，使其到达不变分布，然后令β $\to$ ＋∞，以达到 f 取最小值的点集上．而固定 b 时的 Markov 链的发展的取样过程，则可以通过上一节所叙述的各种 Markov Monte Carlo 方法来实现．例如可用 Gibbs 采样法，Metropolis 采样法，或条件概率的采样法．(8.20)中的不变分布的形式，很像在物理上的 Boltzmann 分布，而 b 则是倒温度，即 b ＝ $\frac{1}{T}$ （T 是绝对温度）．若把无噪声(即 b ＝＋∞情形)的决定性算法，看成是绝对零度(T ＝$\frac{1}{b} = 0$)的情形，那么，β＜∞的情形就是将温度升为正温度．于是上面所述的达到目标函数取最小值的集合的过程，即对固定的 b，构造 Markov 链的样本，再让 $b \to \infty$ 的过程，恰似先将温度由绝对零度升高（即 b ＜∞），然后再让温度降回到绝对零度（即令 b =＋∞）的过程．这个升温再降温的过程，就是“退火”过程的模拟．因而这个算法称为模拟退火算法，或称随机松弛法．简言之，即升温以得到不变分布，再降温(b $\to$ ＋∞)使不变分布集中于目标函数 f 的取最小值的集合上．

为了加快计算速度，人们在理论上发现，可以采取受人工噪声干扰的非时齐的 Markov 链，把 Markov 链的时间发展与降温这两步，归并成一步作模拟退火． 即使所选的 Markov 链的第 n 步转移矩阵随 b_n 而改变，当时间 n 趋于无穷时，只要 b_n (称为冷却进度（Cooling Schedule））以适当缓慢的速度趋于∞ ，则此非时齐的 Markov 链，就依概率收敛到目标函数达到最小值的点集上．这个非时齐的退火的过程完全由冷却进度所控制，如果 b_n 减小过快，则在直观上就相当于在每个温度段还未达到平衡，退火过程就还会陷入能量函数的局部极小；反之，如果 b_n 减小过慢，相当于几乎是常值，退火过程就会按平衡态(即不变分布)发展．这两种情形都不能达到整体极小．

更确切地说，存在一个由目标函数 f 确定的正常数 a， 只要取

$$b_n = g \cdot \log n \qquad (g < a), \qquad (8.21)$$

(即取温度$T=\frac{1}{b_n}$),则当 $n\to\infty$ 的进程就达到了“适当缓慢地降低温度”这个要求．当然g 取得越接近a 效果越好，计算接近目标函数取最小值的点集的进程越快．

在实际计算时，当n大时，按(8. 21)式取的b_n的变化非常小，以致在实际上几乎看不出变化．所以有时人们就简便地使用一个固定的很小的b，构造具有（８．２０）式的不变分布的 Markov 链，用其发展以代替模拟退火．显然，这种 Markov 发展的终极实际上是一个平衡分布．但是只要这个平衡分布高度集中于目标函数全局极小的集合附近，那么这样的计算误差也就可以容忍了．在具体操作时，到底如何掌握分寸，是一个实际经验问题，至今还缺少中肯的理论分析．

在许多情况下，由于所处理的系统非常复杂，或者实际维数太高，还有的时候目标函数没有明确的分析表达式，这时人们就引入局部改变的邻域：记状态i的邻域为$N(i)$（这种邻域是根据问题的要求自然地确定的），作矩阵

$$\tilde{P}=(\tilde{p}_{i,j}),\ \tilde{p}_{i,j}=\begin{cases}|N(i)|^{-1} & (j\in N(i))\\ 0 & (j\notin N(i))\end{cases},\qquad （８．２２）$$

它规定了只许在近邻之间转移，即当系统处于状态i 时，它下一时刻只能转移为$N(i)$中的状态．再规定所要的 Markov 链第 n 步转移概率矩阵由如下的 Metropolis 采样：

$$\mathbf{P}^{(n)} \overset{\Delta}{=} (p_{ij}(b_n)),\qquad （８．２３）$$

$$p_{i,j}(b)=\begin{cases}\tilde{p}_{ij}\cdot\min\{1,\dfrac{p_j^{(b)}}{p_i^{(b)}}\} & \forall j\neq i\\ 1-\sum_{k\neq i}\tilde{p}_{ik}\cdot\min\{1,\dfrac{p_k^{(b)}}{p_i^{(b)}}\} & j=i\end{cases}\qquad （８．２４）$$

所给定．那么，运转这一族转移矩阵$\{\mathbf{P}^{(n)}\}$所决定的非时齐 Markov 链，只要温度$T_n=\frac{1}{b_n}$以合适的速度下降，就收敛到使f 取最小值的点集．

上面指出的算法的局部化思想，是模拟退火算法中为了克服因为状态太多而引起的计算爆炸的有力手段．这里的矩阵$\tilde{P}$可以一般化地取一个互通的概率转移矩阵，显见它越稀疏（即其中 0 越多)则进程可能越快．它称为局部化预选矩阵，简称预选矩阵．它相当于规定了可以转移的“邻域”，把整体搜索变为局部搜索．以减少计算量．在研究第 9 章中的 Gibbs 场的 MCMC 及模拟退火时，就以邻域转移代替这个预选矩阵．

3.2 有关模拟退火算法的非时齐马氏链的理论背景

我们的仍旧假定状态集(或称为组态空间)是有限集， 即$S=\{1,2,\cdots,M\}$．设f为S上的一个实值函数．令 L(S)为 S 上全体实值函数的集合． 对于任意 $f=(f(1),\cdots,f(M))$ ，$g=(g(1),\cdots,g(M))\in$ L(S)， 定义

$$\text{范数} \| f \| \overset{D}{=} \sum_{i=1}^{M} | f(i) |，\text{距离 } d(f,g) = \| f-g \|.$$

那么 L(S)就成为完备的距离空间．显见S上的概率分布都是 L(S)中的元素，而且S上的任意概率分布的序列在此距离空间中的极限仍是一个概率分布．(但是，如果S不是有限集而是具有可数个元素的集合时，此结论不真)．

由定义我们立即推得：对于任意一个转移矩阵$\mathbf{P}$，恒有 $\| f\ \mathbf{P} | \le \| f \|$，其中$f\ \mathbf{P}$指$f$与$\mathbf{P}$作为矩阵相乘．

(1) 时齐的 Markov 链优化的模拟退火方法

定理８．２　设以 $\mathrm{P}(b)=(p_{ij}(b))$为转移矩阵的 Markov 链记为$\{x_n^b\}$，则

$$\lim_{b\to\infty}\lim_{n\to\infty} P(f_{x_n^b}=f_*)=1.$$

证明

$$\lim_{b\to\infty}\lim_{n\to\infty} P(f_{x_n^b}=f_*)=\lim_{b\to\infty}\lim_{n\to\infty} P(x_n^b\in\{i:f(i)=f_*\})$$
$$=\lim_{b\to\infty}\sum_{\{i:f(i)=f_*\}} p_i^{(b)}=1.$$

(2) 非时齐 Markov 链的模拟退火

非时齐 Markov 链的分布的收敛性有

定理８．３　(Dobrushin-Isaacson-Madsen 定理)　设x_n是状态空间S上的非时齐 Markov 链，其第n步转移矩阵为$\mathbf{P}^{(n)}$．如果它们满足：

(A.1) $\mathbf{P}^{(n)}$作为时齐的转移矩阵时有不变分布$p^{(n)}$；

(A.2) $\sum_n \| p^{(n)}-p^{(n+1)} \| < \infty$；

(A.3) 或者满足 Isaacson-Madsen 条件：对于任意概率分布向量m,n，及正整数j，恒有

$$\| (\mathrm{m}-\mathrm{n})\mathbf{P}^{(j)}\cdots\mathbf{P}^{(n)} \| \xrightarrow{n\to\infty} 0;$$

或者满足 Dobrushin 条件：对于任意 j，收缩系数满足

$$C(P^{(j)}\cdots P^{(n)}) \xrightarrow{n\to\infty} 0.$$

那么，

(1) $p^{(\infty)} \overset{\Delta}{=} \lim_{n\to\infty} p^{(n)}$存在；

（2）无论什么初分布 m_0 下，此非时齐的 Markov 链 $\{x_n\}$ 在时刻 n 的绝对分布 m_n 恒有极限

$$m_n \to p^{(\infty)}.$$

证明　首先，由(A.2)利用三角形不等式立得 $\{p^{(n)}\}$ 是 L(S)中的 Cauchy 列，故而(1)正确．往证(2)．用三角形不等式与 Dobrushin 不等式，我们得到

$$\| m_0\mathbf{P}^{(1)}\cdots\mathbf{P}^{(\mathbf{n})} - p^{(\infty)} \|$$

$$\le \| (m_0\mathbf{P}^{(1)}\cdots\mathbf{P}^{(\mathbf{n})} - p^{(\infty)})\mathbf{P}^{(\mathbf{j})}\cdots\mathbf{P}^{(\mathbf{n})} \| + \| p^{(\infty)}\mathbf{P}^{(\mathbf{j})}\cdots\mathbf{P}^{(\mathbf{n})} - p^{(\infty)} \|$$

$$\le 2C(P^{(j)}\cdots P^{(n)}) + \| p^{(\infty)}\mathbf{P}^{(\mathbf{j})}\cdots\mathbf{P}^{(\mathbf{n})} - p^{(\infty)} \|,$$

在满足 Isaacson-Madsen 条件时，上式的第一项趋于 0，(而由 Dobrushin 不等式推出

$$\| (\mathrm{m}_0\mathbf{P}^{(1)}\cdots\mathbf{P}^{(\mathbf{n})} - \mathrm{p}^{(\infty)})\mathbf{P}^{(\mathbf{j})}\cdots\mathbf{P}^{(\mathbf{n})} \| \le 2C(P^{(j)}\cdots P^{(n)}),$$

即在 Dobrushin 条件成立时，Isaacson-Madsen 条件必然成立)．对第二项利用不变分布条件 $p^{(\mathbf{m})} = p^{(\mathbf{m})}\mathbf{P}^{(\mathbf{m})}$ 推得对于 $j \ge M$ 有

$$\| p^{(\infty)}\mathbf{P}^{(\mathbf{j})}\cdots\mathbf{P}^{(\mathbf{n})} - p^{(\infty)} \|$$

$$= \| (p^{(\infty)} - p^{(\mathbf{j})})\mathbf{P}^{(\mathbf{j})}\cdots\mathbf{P}^{(\mathbf{n})} + (p^{(\mathbf{j})}\mathbf{P}^{(\mathbf{j})}\cdots\mathbf{P}^{(\mathbf{n})} - p^{(\infty)}) \|$$

$$= \| (p^{(\infty)} - p^{(\mathbf{j})})\mathbf{P}^{(\mathbf{j})}\cdots\mathbf{P}^{(\mathbf{n})} + (p^{(\mathbf{j})}\mathbf{P}^{(\mathbf{j+1})}\cdots\mathbf{P}^{(\mathbf{n})} - p^{(\infty)}) \|$$

$$= \| (p^{(\infty)} - p^{(\mathbf{j})})\mathbf{P}^{(\mathbf{j})}\cdots\mathbf{P}^{(\mathbf{n})} + \sum_{\mathbf{k}=1}^{\mathbf{n-j}} (p^{(\mathbf{j+k-1})} - p^{(\mathbf{j+k})})\mathbf{P}^{(\mathbf{j+k})}\cdots\mathbf{P}^{(\mathbf{n})} + (p^{(\mathbf{n})} - p^{(\infty)}) \|$$

$$\le \sup_{n\ge M} \| (p^{(\infty)} - p^{(n)}) \| + \sum_{n\ge M} \| p^{(n)} - p^{(n+1)} \| + \sup_{n\ge M} \| p^{(n)} - p^{(\infty)} \| \to 0.$$

［注］　在乘积 $\mathbf{P}^{(\mathbf{j})}\cdots\mathbf{P}^{(\mathbf{n})} \overset{\Delta}{=} (p_{ik}^{(j,n)})_{i,k\le M}$ 中，$p_{ik}^{(j,n+1)} = P(x_{n+1} = k \mid x_j = i)$ 为第 n 次转移矩阵为 $\mathbf{P}^{(n)}$ 的非时齐 Markov 链 $\{x_n\}$ 的多步转移概率．由于我们研究的是有限状态的情形，于是 Isaacson-Madsen 条件可以改写为如下的等价条件(称为弱遍历条件)：

(A. 2)′ 对于任意状态 $i, l, k \in S$，任意正整数 j，恒有

$$\lim_{n\to\infty} (p_{ik}^{(j,n)} - p_{lk}^{(j,n)}) = 0.$$

对于模拟退火的收敛性，利用较为深入的，转移概率为指数型衰减的时齐的 Markov 链序列的大偏差理论（粗略地说，它研究此链越出给定小概率状态集合的概率的指数型衰减)，可以证明(这需要相当深入的叙述与证明，本书无法列入）下面的定理 8.4．

设 $f(i)$ 为定义在 $S = \{1, \cdots, M\}$ 上的函数．令

$$\pi^{(n)} = (p_1^{(n)}, \cdots, p_M^{(n)}),\ p_i^{(n)} = \frac{1}{\sum_j e^{-b_n \cdot f(j)}} e^{-b_n \cdot f(i)}, \tag{8.25}$$

$$\mathbf{P}^{(n)} \stackrel{\Delta}{=} (p_{ij}^{(n)}),\ p_{ij}^{(n)} = \begin{cases} \tilde{p}_{ij} \cdot \min\{1, \frac{p_j^{(n)}}{p_i^{(n)}}\} & \forall j \neq i \\ 1 - \sum_{k \neq i} \tilde{p}_{ik} \cdot \min\{1, \frac{p_k^{(n)}}{p_i^{(n)}}\} & j = i \end{cases}. \tag{8.26}$$

定理8.4　(Gemam-Gemam)

假定

(1) $0 < b_n \uparrow \infty$，满足：存在n_0，使$n \ge n_0$后，有

$$b_n \le \frac{1}{M\Delta} \ln n, \tag{8.27}$$

其中Δ是目标函数f的振幅：

$$\Delta \stackrel{\Delta}{=} \max_i f(i) - \min_i f(i); \tag{8.28}$$

(2) 预选矩阵$\tilde{\mathbf{P}} = (\tilde{p}_{ij})$是互通的概率转移矩阵.

那么，以$\mathbf{P}^{(n)}$为第n步转移矩阵列的非时齐的 Markov 链$\{x_n\}$满足弱遍历条件，因而(A. 2)满足.

定理8.5　(模拟退火的收敛性定理)

在定理 8.4 的条件满足时，无论什么初始分布μ_0下，以$\mathbf{P}^{(n)}$为第n步转移矩阵列的非时齐的 Markov 链$\{x_n\}$，在时刻n的绝对分布μ_n有极限$\mu_n \to \pi^{(\infty)}$，其中$\pi^{(\infty)}$为集合$\{i: f(i) = \min\}$上的离散均匀分布．从而有

$$P(f_{x_n} \xrightarrow{n \to \infty} f_*) = 1.$$

证明　我们来验证 Dobrushin-Isaacson-Madsen 定理的条件成立．由 Gemam-Gemam 定理，我们只需验证条件(A. 1)． 为此，首先注意当n充分大时(例如$n > n_0$时)，利用b_n的单调不减性，我们有：对于固定的i，$p_i^{(n)} = \frac{1}{\sum_j e^{-b_n \cdot f(j)}} e^{-b_n \cdot f(i)}$ 在$f(i) > 0$时，关于n是单调不减的(对b_n的微商非负)；而在时$f(i) \le 0$ 时，关于n是单调不增的(对b_n的微商非正)．再注意(这是在第 5 章 Dobrushin 不等式中应用过的事实)

$$\|\pi^{(n+1)} - \pi^{(n)}\| = \sum_i \|p_i^{(n+1)} - p_i^{(n)}\| = \sum_i (p_i^{(n+1)} - p_i^{(n)})^+ + \sum_i (p_i^{(n+1)} - p_i^{(n)})^-$$
$$= 2\sum_i (p_i^{(n+1)} - p_i^{(n)})^+ = 2\sum_i (p_i^{(n+1)} - p_i^{(n)})^-.$$

于是我们得到

$$\sum_{n=n_0}^{N}\|\mathrm{p}^{(n+1)}-\mathrm{p}^{(n)}\|=2\sum_{n=n_0}^{N}\sum_{i}(p_i^{(n+1)}-p_i^{(n)})^+=2\sum_{f(i)>0}\sum_{n=n_0}^{N}(p_i^{(n+1)}-p_i^{(n)})^+$$

$$=2\sum_{f(i)>0}\sum_{n=n_0}^{N}(p_i^{(n+1)}-p_i^{(n)})=2\sum_{f(i)>0}[p_i^{(N+1)}-p_i^{(n_0)}]\le 2.$$

令 $N\to\infty$，便得 $\sum_{n=n_0}^{\infty}\|\mathrm{p}^{(n+1)}-\mathrm{p}^{(n)}\|\le 2$. 所以 (A. 1)成立. 定理证毕.

习题8

1. 证明对于 Metropolis 采样在时刻 n 的更新 $x^{(n)}\to x^{(n+1)}$ 采取的操作得到的由 i 到 j 的转移,就是（８．１５）式定义的转移概率.

2. 证明由（８．１６）给出的转移矩阵所对应的 Markov 链以 p 为可逆分布. 再给出这种 Metropolis 采样在时刻 n 的更新 $x^{(n)}\to x^{(n+1)}$ 可具体采取如下的操作步骤.

3. 证明由 Hastings 推广了的 Metropolis 采样所对应的 Markov 链以 p 为可逆分布. 再给出这种采样在时刻 n 的更新 $x^{(n)}\to x^{(n+1)}$ 可具体采取如下的操作步骤.

龚光鲁，钱敏平著 应用随机过程教程及其在算法与智能计算中的应用
清华大学出版社, 2003

第 9 章 以图象信息为背景的随机场, 迭代 Markov 系统

随机场的一个直观背景，是以图像为基本事件的随机元素．另一方面，迭代 Markov 系统是取值为压缩映射随机映射的迭代，其目的是得到其不变元素，以用来描述一个图像.

1 有限格点上的 Markov 随机场与图象

1. 1 有限格点上的 Markov 随机场

定义９.１（相邻系统与子团）

设 G 为平面有限格点集合：$G=\{(i,j):1\le i\le N_1, 1\le j\le N_2\}$．对于 G 中的任意两点 $x\neq y$，可以定义它们相邻，或者不相邻．这样定义了 S 中的一种相邻概念以后，就得到了 G 上的一个相邻系统．注意，两点是否相邻完全可以是抽象的，并不需要在直观上比邻(有时为了看起来有对称性，也可以认为 N_1 与 1 恒同，N_2 也与 1 恒同，即认为 G 为环面格点.)

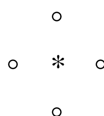
在本书中，我们把 x,y 相邻记为 $x\overset{\partial}{\sim}y$，记与 x 的相邻的点的全体为 $\partial(x)$．G 的子集 C 称为 G 的 $\partial-$ 子团，如果 C 中的任意两个元素都相邻． 含 x 的子团记为 $C(x)=\bigcup\partial(x)$．

一个相邻系统由 G 中点对的相邻概念所决定，因此也由全体子团所决定，所以人们常用 ∂ 来代表一个相邻系统，或用全体子团 $\{C\}$ 来代表一个相邻系统．

例９．２ (1) 若 $\partial(x)=f$，则只有单点集是 G 的 $\partial-$ 子团．

(2) 若 $\partial(x)=G\setminus\{x\}$，则 G 的所有子集都是 $\partial-$ 子团..

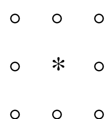
(3) 一阶相邻(最紧相邻，简称紧邻)： ＊ 的相邻点◦可以如下图



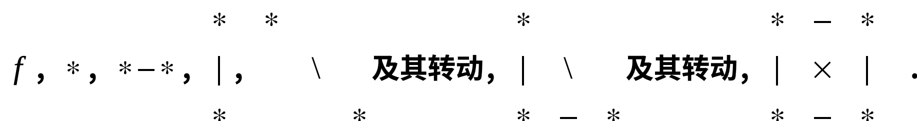
此时 G 的全部 $\partial-$ 子团形式为：

$$f,\ *,\ *-*,\ \begin{matrix}*\\|*\end{matrix}$$

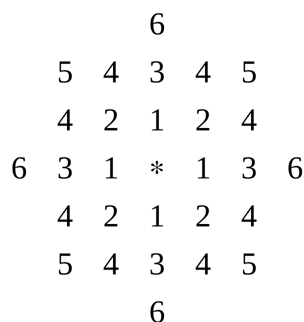
(4)二阶相邻： ＊ 的相邻点◦可以如下图



此时G的全部$\partial-$子团形式为：



高阶相邻等可以如下图所示



其中数字代表分阶为相邻的阶．S的一个相邻系统也就是以G的点为顶点的一个双向图．

定义９．３ （随机场）

以平面矩形格点集G中的格点为参数，取值于$\{0,\cdots,L-1\}$ （也可以是实直线R，或一个离散的集合)的随机变量族，称为取值于$\{0,\cdots,L-1\}$的$G-$随机场，简称为随机场．按照统计物理中的术语，我们称$\{0,\cdots,L-1\}$为自旋空间．

于是，随机场的一个样本，就是在格点集G的每个格点上 ”坐着” $\{0,\cdots,L-1\}$中某一个数．随机场的样本的一个直观背景是图像．黑白图象是连续变化的一个灰度图．而在实际图象处理中的黑白灰度都是经过离散化采样的．设黑白灰度可分成L个 (一般 L=256)不同的等级．$\{0,\cdots,L-1\}$中的一个数值可以解释为黑白的灰度．最粗放的是黑白轮廓图，在每个采样点上只有 ”黑”, “白” 两个象素之一 (即$L=1$的情形)． 一般G为256×256个采样点 (即$N_1=N_2=256$)．那么，一幅黑白灰度图就可以看成在G的每一个格点上的一个灰度(象素,Pixel)，即给每一个格点赋予$0,\cdots,L-1$中的一个值．所以，$G-$随机场的样本，即做一次试验可能得到的结果，可以解释为G上的一个黑白灰度图．

在这个意义上，随机场就也可以解释为随机图，或者说，它是在图像空间取值的一个随机元素(简称为随机元)．这是随机场的第 2 种解释．从相对灰度(每个格点上的灰度除以总灰度(灰度的和))的角度也可以说，随机场的分布是格点集G上的一个随机的概率分布．

定义９．４（组态） 从数学的角度抽象，随机场的一个样本就是$S\overset{\Delta}{=}\{0,\cdots,L-1\}^G$的一个元素，记为$a,b\cdots$，称为一个组态(configuration)，即$a=(a_x:x\in G)$，

$a_x \in \{0,\cdots,L-1\}$称为组态a的x坐标．再记$a_J=(a_x: x\in J(\subset G))$，称为组态$a$的$J$坐标．

所以，随机场也可以看成是一个随机组态，或者看成为在组态空间$S=\{0,\cdots,L-1\}^G$上取值的随机元．这是随机场的第 3 种解释．从分布的角度也可以说，随机场的分布是组态空间上的一个概率分布．

由此可见，一个灰度图就可以近似地视为一个组态．

定义９．５　（$\partial-$Markov 随机场）

随机场$\{x_x: x\in G\}$称为$\partial-$Markov 场，如果对于任意格点$x,y\in G$及任意组态$a\in S=\{1,\cdots,L\}^G$，恒有

$$P(x_x=a_x \mid x_y=a_y,\text{一切}y\neq x)=P(x_x=a_x \mid x_y=a_y,\text{一切}y\overset{\partial}{\sim}x).\quad (9.1)$$

$\partial-$Markov 场的含义是：在任意格点x的余格点位置上取已知值的条件下，随机场在格点x处取值的概率规律只与格点x的∂相邻点有关．

由定义可以直接证明下述定理

定理９．６　随机场$\{x_x: x\in G\}$是$\partial-$Markov 场的等价条件为：　对任意格点集$A\subset G$及任意组态$a\in S=\{0,\cdots,L-1\}^G$，恒有

$$P(x_x=a_x, x\in A \mid x_y=a_y, y\notin A)=P(x_x=a_x, x\in A \mid x_y=a_y, y\in\partial(A)),$$

其中　$\partial(A)\overset{\Delta}{=}\{A\text{中点的邻点全体}\}$．

1．２　相邻系统的 Gibbs 分布与 Gibbs 随机场（$\partial-$邻位势 Gibbs 场）

定义９．７（格点集G的子集与组态的相互作用）

对于给定的一个子格点集$A(\subset G)$，组态上的一个与之相关的函数$U_A(a)$（$a\in\{0,\cdots,L-1\}^G$）称为 A 对于组态a的相互作用，如果它满足：　对于任意两个组态$a,b\in\{0,\cdots,L-1\}^G$，只要 $x_x(a)=x_x(b)$（$\forall x\in A$），就有 $U_A(a)=U_A(b)$（$U_A(a)$代表组态a在A处的"坐标"对a的作用，所以要求对在A处"坐标"相同的任意组态的相互作用一样）．

定义９．８（位势函数）　在补充定义$U_f=0$后，相互作用

$$\{U_A: A\subset G\}$$

的全体称为位势．对于给定的相邻系统∂，如果对于任意A，只要它不是$\partial-$子团，就有$U_A=0$，那么 $\{U_C: C\text{为}G\text{的}\partial-\text{子团}\}$就称为$\partial-$邻位势．如果$\partial-$相邻正好就是直观上的紧邻，则这样的的$\partial-$邻位势简称紧邻位势．

定义９.９（能量函数）　对于$\partial-$邻位势，组态的函数

$$H_U \stackrel{\Delta}{=} \sum_{\partial\text{子团}C} U_C \qquad (9.2)$$

称为$\partial-$邻位势的能量函数.

定义９.１０（$\partial-$邻位势的 Gibbs 分布与 Gibbs 随机场）

在组态空间$S=\{0,\cdots,L-1\}^G$上定义的概率分布$p=p(a),(a\in S)$，称为$\partial-$邻位势(在能量函数下)的 Gibbs 分布(这里是离散分布)，其中

$$p(a)=\frac{1}{Z}e^{-H_U(a)},$$

H_U为$\partial-$邻位势的能量函数，而 Z 为一个归一化常数，称为 Gibbs 分布的配分常数：

$$Z=\sum_{a\in S}e^{-H_U(a)}. \qquad (9.3)$$

以 Gibbs 分布为分布的随机场称为 Gibbs 场，即

$$P(x_x=a_x,\forall x\in G)=p(a).$$

就是说，Gibbs 分布 p 是随机场在所有的格点处取值的联合分布.

在第 6,7 章中的 Ising 模型中的 Gibbs 分布，就是这里的 Gibbs 场的概率分布的特殊情形，即$L=2$，自旋空间为$\{-1,1\}$的情形，其对应的位势是紧邻的(而把外场看成格点对自己的作用的和). Ising 模型中 Gibbs 分布中有一个倒温度参数，这是为了调节用的．这里无妨省去.

再则，由于在不同的邻域中的科学家有他们习惯用的记号，所以，在本节中的记号与第 6,7 章中用的有些不同，请读者见谅由于记号改动所带来的不便．我们列出其对比：

	组态	配称分布	倒温度	随机场
第6,7章	x,h	a,p	b	X_t
本节	a,b	p		x_t

在第 6,7 章中，由于 Ising 模型较为简单，我们只提出了 Gibbs 分布的概念，而并未提出随机场的概念．事实上，Gibbs 分布是 Gibbs 随机场的不变分布（ 有的文献中 Gibbs 分布与 Gibbs 随机场的术语是混用的）．因此在第 6,7 章中，我们如很多书上那样用x,h表示组态.

而在本节中，我们强调的是随机场，它用x,h表示更为合适，所以我们就与很多书上不一样，

改用a,b表示组态.

定理 9.11　(Hammersley-Clifford 基本定理)

对于给定的相邻系统∂，$\partial-$Markov 随机场与$\partial-$邻位势 Gibbs 场是一样的.　】

这个定理的意义在于，用$\partial-$Markov 随机场来描述图象是一个直观认识，而用$\partial-$邻位势 Gibbs 场则可归结为其能量函数，后者有时可以用简单的参数来表达(例如相互作用具有平移不变性的时候)．那时就可用几个参数来描述图象．从而可以达到信息压缩的目的．于

是，此定理将直观认识转化为参数表示，以便于数学处理．

１．３　图象处理的随机过程方法的思路原则概述

正如前面所述，一幅黑白图象是某个组态空间中的一个组态．在图像传输过程中，需要对于灰度图的信息进行处理与压缩．对于接收到的信息还需要恢复（Retrieving）为原图像．常见的方法有：

第 1 种方法（用能量函数的 Gibbs 分布对图像建模）

按黑白图像的相对灰度(即每个格点的灰度除以图像的总灰度和)大小，把图像看成格点集 G 上的组态空间 $S=\{0,1,\cdots,L\}^G$ 中的一个组态，或者看成 G 上的一个概率分布．由于自然景物等的图象往往具有一些局部相联系(空间不变性)的结构，这些局部结构可以抽象地归结为组态间按某种相邻系统的能量函数．从而可以把一个图像用一个能量函数取到最小值的组态，也就是用 Gibbs 分布达到最大值处的组态，从而可以通过能量函数或 Gibbs 分布来对图像建模． 这个方法的要害处是：如何由图像给出其能量函数．能量函数表示图像能起一定的信息压缩作用．而反过来由能量函数恢复图像的问题，则可以用模拟退火或其它 Markov Monte Carlo 方法运作．

第 2 种方法（用低阶 Markov 场的样本对图像建模）

将黑白图像看成随机场的一个样本．由定理９．１１知道任意随机场都可以看成 Markov 场．而图像的局部结构又说明,用较低阶相邻的 Markov 场的样本来描述图象更为合理．

第 3 种方法 (用迭代随机映射的不动点对图像建模)

对于一个黑白轮廓图像，构造与之相系的有限个压缩映射，使得由此压缩映射组给出的一个迭代映射的不动点近似地是此轮廓图．而对于一个黑白灰度图，构造与之相系的一个随机地取有限个压缩映射的随机映射，使此图像近似地是此随机映射的不变分布．（这在第３节中将较多地介绍）．这种方法实际上是处理图像的分形方法的另一角度的发展．

一般地，彩色图可以看成红，黄，兰三张灰度图的叠置．

收到了不清晰的图像，也就是受到了干扰的图像， 就需要对图像进行滤波，称为图像的清洗．其中常见的一种清洗方法为 Bayes 方法，即把未清洗的图像看成先验分布，而寻找转移机制，使清洗后的恢复图象是此先验分布的 Bayes 分布(后验分布)．图像受到污染的典型情形，例如，被测量的图象象素 a_x 受到检测与录制系统的干扰，使实际测到的象素为

$$b_x = g\Big(\sum_{y\in\partial(x)} h_y a_x\Big) + w_x ,$$

其中 g 为一个已知函数， h_y 为已知的通道系数，而噪声 w_x 的统计规律为已知．这时要恢复图象象素 a_x 就需要进行滤波．

此外，图象边缘信息也可能受到干扰而产生变形，也需要用边缘检测来恢复图象的信息．因此有时应在图象的先验信息中加进“边-随机场”的先验分布．

有时在处理图像前可能还需要对图像进行有部分相重叠的分割 (segmentation).

１．４　Gibbs 分布的样本的 Gibbs 采样法(Gibbs Sampler)

要对 Gibbs 场作一些统计估计(例如参数估计)，就先要得到 Gibbs 分布的配分常数．但是由于组态空间 $S=\{1,\cdots,L\}^G$ 的元素数通常非常庞大，例如可以有 10^{10^6} ，这就使直接计算

配分常数 Z 时的求和项太多，从而常会出现大量被求和的项的值小于计算机误差的精度，以致使计算在实际上失去可能．用 Markov 链 Monte Carlo 方法 (MCMC)可以解决这个困难，即先近似地得到 Gibbs 分布，再进行参数估计．所以 MCMC 是一个重要的计算手段.

MCMC 方法就是构造以 Gibbs 分布为可逆分布的 Markov 链的转移矩阵．因为这里的随机场的 Gibbs 分布远比 Ising 模型的 Gibbs 分布复杂，我们只构造离散时间的 Markov 转移矩阵，而且还希望如Glauber动力学那样，每次只在一个格点上改变(称为按Gibbs方式转移)，最后还要保证由这样得到的转移矩阵的各行都收敛于 Gibbs 分布.

为了使构造更为透明，我们先对任给的格点子集 J 定义 如下的"部分转移"

$$p_J(a,b)=\begin{cases}\dfrac{1}{Z_J(a)}e^{-H_U(b_Ja_{S\setminus J})} & (a,b\text{在}J\text{外坐标相同})\\ 0 & (\text{其它})\end{cases},\qquad (9.4)$$

其中 $b_J a_{S\setminus J}$ 表示：把组态 a 中在 J 处的坐标换成组态 b 中的相应坐标，而 $Z_J(a)$ 是一个归一化常数("局部的"配分常数)：

$$Z_J(a)=\sum_{b_J}e^{-H_U(b_Ja_{S\setminus J})}\ .\qquad (9.5)$$

这种 "部分转移" 表示组态间只在位置 J 外的坐标相同时才可能转移.

在此基础上我们定义转移矩阵如下：记 G 中的全部格点的一个给定的排序为 $\{x_1,\cdots,x_{|G|}\},(|G|=N_1N_2)$．任给组态 $a,b\in S$，定义

$$\mathbf{P}_J=(p_J(a,b))_{a,b\in S}\ ,\ \mathbf{P}=\prod_{k=1}^{|G|}\mathbf{P}_{\{x_k\}},\qquad (9.6)$$

则 $\mathbf{P}$ 是一个转移矩阵．注意由于 $\mathbf{P}_{\{x_k\}}$ 中的转移只在一个格点 x_i 上进行，由 $\mathbf{P}$ 构造的Markov 链的一次转移,可以视为 $|G|$ 次连续转移的结果，其中每次转移只在一个格点上进行，而且要做遍所有的格点.

下面的遍历收敛定理是 Markov 链的 Monte Carlo 方法的理论基础．而 Markov 链 Monte Carlo 方法主要用于获得 Gibbs 分布的样本，以便进一步得到 Gibbs 场的各种统计平均量.

对于任意格点 $x\in G$，任意组态 $a\in S$，记

$$H_U(g_xa_{G\setminus\{x\}})\overset{\Delta}{=}m_x\overset{\Delta}{=}\min_{h_x}H_U(h_xa_{G\setminus\{x\}})\ .\qquad (9.7)$$

那么

$$p_{\{x\}}(a,b)=\frac{e^{-[H_U(b_xa_{G\setminus\{x\}})-m_x]}}{\sum_{y\in G}e^{-[H_U(b_ya_{G\setminus\{y\}})-m_x]}}$$

$$\geq\frac{e^{-\max(H_U(b)-H_U(g):b_{G\setminus\{x\}}=g_{G\setminus\{x\}})}}{|G|}\overset{\Delta}{=}\frac{1}{|G|}e^{-d_x}\geq\frac{1}{|G|}e^{-\Delta},$$

其中 d_x 为能量函数 H_U 在格点 x 上的振幅，Δ 为 H_U 的振幅，其确切定义为

$$d_x = \max_{a_{G\setminus\{x\}}=b_{G\setminus\{x\}}} |H_U(a) - H_U(b)|, \quad \Delta \stackrel{\Delta}{=} \max_{x\in G} d_x . \tag{9．8}$$

由此推出

$$C(P_{\{x\}}) \le 1 - |G| \min_{a,b} p_{\{x\}}(a,b) \le 1 - e^{-\Delta} . \tag{9．9}$$

从而得到

$$C(P) = C(P_{\{x_1\}}) \cdots C(P_{\{x_M\}}) \le (1 - e^{-\Delta})^{|G|} \le 1 - e^{-\Delta|G|} . \tag{9．10}$$

定理 9．１２ （遍历收敛定理） 以上定义的转移矩阵 $\mathbf{P}$ 的 Dobrushin 收缩系数满足

$$C(P) \le 1 - e^{-\Delta|G|} \quad (\Delta \text{为} H_U \text{的振幅}).$$

记以 $\mathbf{P}$ 为转移矩阵的 Markov 链为 x_n，则它以 Gibbs 分布 p 为可逆分布，而且有

$$\mathbf{P}^n \xrightarrow{n\to\infty} \mathbf{1}^T p .$$

证明 显见 $p(a)p_J(a,b) = p(b)p_J(b,a)$．由 $\mathbf{P} = (p(a,b))_{a,b\in S}$ 的定义立得 $p(a)p(a,b) = p(b)p(b,a)$．因此 p 为可逆分布．又由 $C(P) < 1$，用 Dobrushin 定理可知极限成立．

［注］ G-格点的排序 $\{x_1, \cdots, x_{|G|}\}$ 并不是实质的．我们可以用一个 G 上取值全正的预选概率分布 $g(x)$，(即 $g(x) > 0, \sum_{x\in G} g(x) = 1$ 例如均匀分布）的随机数作随机选取来代替．这时转移矩阵 $\mathbf{P}$ 应作相应的改动，代替 $p_{\{x_k\}}(a,b)$ 的是，以 $g(x)$ 的概率对只在 x 坐标不同的组态间进行转移，即令

$$\mathbf{P} \stackrel{\Delta}{=} (\tilde{\mathbf{P}})^{|G|}, \quad \tilde{\mathbf{P}} = (\tilde{p}(a,b))_{a,b\in S} ,$$

其中

$$\tilde{p}(a,b) = \begin{cases} g(x)p_{\{x\}}(a,b) & (\text{若} a_{G\setminus\{x\}} = b_{G\setminus\{x\}}) \\ 0 & (\text{其它情形}) \end{cases} .$$

这时我们仍有

$$\mathbf{P}^n \xrightarrow{n\to\infty} \mathbf{1}^T p .$$

１．５ Gibbs 分布的模拟退火

本段讨论求得达到组态空间上能量函数最小值的组态的概率方法，就是用能量函数的 Gibbs 分布作模拟退火的方法．考虑依赖于温度 T_n 的 Gibbs 分布，以它们作为不变分布构造各步为非时齐 Gibbs 转移矩阵，由此确定的一个非时齐 Markov 链 $\{x_n\}$．我们可以用它作模拟退火，使得当 n 充分大以后，此非时齐的 Markov 链 x_n 以大概率落在能量函数最小的组态

所组成的集合中.

更具体地说，对于如下的依赖于温度T_n的$\partial-$邻位势 Gibbs 分布

$$p^{(n)}(a)=\frac{1}{Z_n}e^{-\frac{1}{T_n}H_U(a)}\quad,\quad Z_n=\sum_{a\in S}e^{-\frac{1}{T_n}H_U(a)}\ ,$$

相应地定义

$$p_J^{(n)}(a,b)=\begin{cases}\dfrac{1}{Z_{n,J}(a)}e^{-\frac{1}{T_n}H_U(b_J a_{S\setminus J})} & (a,b\text{在}J\text{外坐标相同}) \\ 0 & (\text{其它})\end{cases}\ ,\qquad (9.11)$$

$$Z_{n,J}(a)=\sum_{b_J}e^{-\frac{1}{T_n}H_U(b_J a_{S\setminus J})}\ .$$

再令

$$\mathbf{P}_{\{x\}}^{(n)}\overset{D}{=}(=(p_{\{x\}}^{(n)}(a,b))_{a,b\in S}\ ,\quad \mathbf{P}^{(\mathbf{n})}=\prod_{x\in G}\mathbf{P}_{\{x\}}^{(n)}\ ,\qquad (9.12)$$

则$\mathbf{P}^{(n)}$是一个转移矩阵. 我们将以它为非时齐转移矩阵列所确定的非时齐 Markov 链记为

$\{x_n\}$.

定理９.１３ (Gibbs 分布的模拟退火的收敛性定理)

如果温度T_n满足如下的 Geman-Geman 条件： 对$0<T_n\downarrow 0$，存在n_0使$n\geq n_0$后有

$$T_n\geq\frac{|G|\Delta}{\ln n}\ ,\qquad (9.13)$$

其中D 是能量函数$H_U(a)$的振幅：

$$D\overset{D}{=}\max_{a\in S}H_U(a)-\min_{a\in S}H(a)\ ,$$

那么在任意初始分布m_0下，以$\mathbf{P}^{(n)}$为第n步转移矩阵列的非时齐的 Markov 链$\{x_n\}$，在时刻n的绝对分布m_n有极限

$$m_n\to p^{(\infty)},$$

其中$p^{(\infty)}$为集合$\{a\in S:H_U(a)=\min\}$上的离散均匀分布. 从而还有

$$\lim_{n\to\infty}P(H_U(x_n)=\min_{a\in S}H_U(a))=1\ .$$

证明　我们验证 Dobrushin-Isaacson-Madsen 定理的条件成立. 条件(A.1)的验证与第 8 章 3. 2 段的模拟退火定理中的证明完全一样. 下面验证 Dobrushin 条件(A. 2)′. 这里定理９.１２中的不等式变为

$$C(P^{(n)}) \le 1 - e^{-\frac{|G|}{T_n}\Delta},$$

其中Δ为能量函数H_U的振幅．于是对于 $\mathbf{P}^{(j)}\mathbf{P}^{(j+1)}\cdots\mathbf{P}^{(n)}$，我们有

$$C(P^{(j)}P^{(j+1)}\cdots P^{(n)}) \le C(P^{(j)})C(P^{(j+1)})\cdots C(P^{(n)}) \le \prod_{k=j}^{|G|}(1-e^{-\frac{|G|\Delta}{T_n}}).$$

由无穷乘积的理论， 当$n\to\infty$时，上面右方趋于 0 的充要条件为级数$\sum_n e^{-\frac{|G|\Delta}{T_n}}$是发散的．但是由 Gemam-Gemam 条件，我们有

$$\sum_n e^{-\frac{|G|\Delta}{T_n}} \ge \sum_n e^{-\ln n} = \sum_n \frac{1}{n} = \infty.$$

所以 Dobrushin 条件(A. 2)'成立.

［注 1］ Gibbs 场的模拟退火在图象的清污(图象的滤波)中的应用

设图象为组态a，而在传输过程这受到了污染，使得接受方观测到的图象为组态b．需要用数学方法进行滤波，以便恢复图象的清晰度．我们把在观测为组态b的条件下，状态a的分布，记为$P(a\mid b)$ (就是后验分布 Posteriori distribution，即 Bayes 分布)．一般它也可用写成 Gibbs 分布的形式:

$$P(a\mid b) = \frac{1}{Z(b)}e^{-H(a|b)}, \qquad (9.\ 14)$$

其中$H(a\mid b)$称为条件能量函数．典型的例子如

$$H(a\mid b) = q\sum_{x\in\partial(y)}(a(x)-a(y))^2 + \frac{1}{2s^2}\sum_x (a(x)-b(x))^2,$$

其中第一项强调了图形的光滑性，第二项为保真度．滤波的目的就是要找组态$\hat{a}$，使条件能量函数达到最小:

$$H(\hat{a}\mid b) = \min_a H(a\mid b).$$

这个最佳组态$\hat{a}$可以通过模拟退火方法求得.

又例如，在边缘检测中，b就是离散化点上的象素，而a就是“边随机链”的一个样值，与滤波问题的不同的是，此时的后验图形只是先验图形的一部分，即边的位置．在复杂的图形处理技术中，例如，X 断层摄影术的处理，情况就远为复杂． 适当的利用分割技术可以得到较好的恢复效果．

［注 2］图象的处理有许多数学方法，Gibbs 分布方法只是其中的一种．另一种更为常用的方法是，把图象的一些特征参数作为观察量(这同时也达到了图象信息压缩的目的．用它们来估计 Markov 链的状态 ，就是重建图象．这就是隐 Markov 方法(HMM)，一般地说，如果特征参数取得合适，HMM 方法是十分有效的．例如，纹理图象分析方法(Texture Analysis)提取特征参数．其它的方法还有：下节将介绍的随机函数迭

代系统方法．除此以外，还有小波分析等非概率方法．

［注 3］ 当格点集 G 为无限集时，组态的数目就不再是可数的了．此时就应归入连续状态的 Markov 场的范畴．在统计物理中应用的模型，就是这种模型．有时自旋空间 S 也可以是连续的，例如，圆周或其它集合．这时就可能存在多个不变分布，称为有相变，因为每个分布代表了统计物理中质点的一个宏观“相”．统计物理中也正是对这种相变的存在性更感兴趣．

［注 4］若每次发展的修改的格点集合 J 不是单点集，且各次的修改集可以不同，那么，此时的 Markov 链未必有遍历定理，这时并不能保证此 Markov 链 x_n 的时间平均收敛到不变分布，因此，即使 n 很大，也不能认为 x_n 的样本近似为不变分布的样本．

2　时间离散状态连续的 Markov 链

2. 1　概率空间再访

设 $\Omega=\{w: w-$基本事件$\}$，F 是样本空间 Ω 的某些子集组成的集合类(事件体)，满足

$$\text{对于任意 } A, A_n \in \text{F} \quad (n \geq 1)\text{，恒有 } \bigcup_{n=1}^{\infty} A_n, \bigcap_{n=1}^{\infty} A_n, \overline{A} \overset{D}{=} W - A \in \text{F}.$$

(这样的事件体 F 代表了可以通过理论推算，或实际测量能够得到概率的全体随机事件)．在概率理论中 F 称为 Ω 上的一个 $s-$代数．(Ω, F) 称为概率空间．于是概率 P 就是定义在 $s-$代数 F 上的一个非负函数．

定义９．１４ (Borel 集与 Borel 函数)

设 B 是 R^d 的某些子集组成的一个 $s-$代数，且满足以下条件：

(1) R^d 中的任意开集 $G \in$ B，

(2) 任意一个有 R^d 的某些子集组成的 $s-$代数 G，只要它含有一切开集，则就有 B $\subset$ G．

那么 B 称为 R^d 上 Borel 集合类，其中的集合称为 Borel 集．即 Borel 集合类是以开集为元素的最小 $s-$代数．(以前在本书中的所谓 “常见的” 集合，均指 Borel 集)．R^1 上的 Borel 集称为一维 Borel 集．

再设 M 为 R^d 上的一个满足下述条件的函数类：

(1) 对极限封闭：若 $f_n \in$ M 且 $f_n(x) \to f(x)$，则有 $f \in$ M，

(2) 任意连续函数 $f \in$ M，

(3) 若 R^d 上一个函数类 L，只要对极限封闭，而且包含全体连续函数，则有 M $\subset$ L．

那么 M 称为 R^d 上 Borel 函数类，其中的函数称为 R^d 上的 Borel 函数.（以前在本书中的所谓“常见的” 函数，均指 Borel 函数）.

再则，随机变量的概念是依赖于事件体(s −代数)F 的，严格地应称为 F-随机变量.

定义9.15 样本空间Ω上的实值函数x 称为 F-可测的，如果对于任意实数a，恒有$\{w : x(w) \le a\} \in$ F. 于是，x 为 F-随机变量与x 为 F-可测的是一样的. 而随机变量的分布则依赖于概率 P 的给法.

分布函数为 $F(x)$ 的随机变量x 的函数 $g(x)$ 的数学期望是如下的 Stieltjes 积分

$$Eg(x) = \int g(x)dF(x) \ (\overset{def}{=} \int g(x)F(dx) \ )$$

（括号中的 $F(dx)$ 就是 $dF(x)$，前一种记号便于推广到多变量，或条件分布函数的情形）.

而

$$F(x\,|\,h) \overset{\Delta}{=} P(x \le x\,|\,h) \overset{\Delta}{=} E(I_{(-\infty,x]}(x)\,|\,h) = [E(I_{(-\infty,x]}(x)\,|\,h = y)]_{y=h}$$

称为x 对于h 的条件分布函数. 类似地，我们有

$$E[g(x)\,|\,h] = \int g(y)F(dy\,|\,h) \,.$$

（注意，右方的积分是直观地理解的，如果想要一个清晰的证明，在数学上还必须要作一系列 “后台操作”）.

2.2 时间离散状态连续的 Markov 链

定义9.16 实值随机序列$\{x_n : n \ge 0\}$称为（时间离散状态连续的）Markov 链，如果对于任意 $n \ge 0$，任意实数 $y, x, x_0, \cdots, x_{n-1}$，有

$$P(x_{n+1} \le y\,|\,x_n = x, x_{n-1} = x_{n-1}, \cdots, x_0 = x_0) = P(x_{n+1} \le y\,|\,x_n = x) \,. \qquad (9.15)$$

（这正说明条件分布函数有 Markov 性）. 在$(x_{n+1}, x_n, \cdots, x_0)$有联合密度时，(9.15)等价于$x_{n+1}$的条件分布密度的如下的等式

$$p_{x_{n+1}}(y\,|\,x_n = x, x_{n-1} = x_{n-1}, \cdots, x_0 = x_0) = p_{x_{n+1}}(y\,|\,x_n = x) \,. \qquad (9.15)'$$

与第5章中类似地，这里的 Markov 性可以有如下的多种等价性描述：

Markov 性的等价性质 1

对于任意实 Borel 集 $L, L_0, \cdots, L_{n-1}$，以及随机事件 $A = \{x_0 \in L_0, \cdots, x_{n-1} \in L_{n-1}\}$，恒有

$$P(x_{n+1} \in L\,|\,A, x_n = x) = P(x_{n+1} \in L\,|\,x_n = x) \,. \qquad (9.16)$$

当存在联合密度时, (9.16)**式也有其相应的条件分布密度形式**.

Markov **性的等价性质** 2

对于随机序列$\{x_n\}$ **在大于时刻** n **所确定的随机事件** $B=\{x_{n+1}\in L_{n+1},\cdots,x_{n+k}\in L_{n+k}\}$

及随机事件 $A=\{x_0\in L_0,\cdots,x_{n-1}\in L_{n-1}\}$ **，有**

$$P(B|A,x_n=x)=P(B|x_n=x) \tag{9. 17}$$

或

$$P(BA)=P(B\,|\,x_n=x)P(A|x_n=x)\,. \tag{9. 17)'}$$

（直观证明　我们只在一切条件分布密度都存在情形证明(9.17)**式**. **先看** $k=2$ **情形**. **由条件概率的性质**, **推广了的全概率公式与** Markov **性的定义**, **我们有**

$$P(B|A,x_n=x)=P(\,x_{n+1}\in L_{n+1},x_{n+2}\in L_{n+2}\,|\,x_n=x,A)$$

$$\text{（直观地）}=P(x_{n+2}\in L_{n+2}\,|\,x_{n+1}\in L_{n+1},x_n=x,A)\;P(x_{n+1}\in L_{n+1}\,|\,x_n=x,A)$$

$$=\int_{\Lambda_{n+1}}P(x_{n+2}\in\Lambda_{n+2}\,|\,x_{n+1}=y,x_n=x,A)f_{n+1}(y)dy\;\;P(x_{n+1}\in\Lambda_{n+1}\,|\,x_n=x)$$

$$=\int_{\Lambda_{n+1}}P(x_{n+2}\in\Lambda_{n+2}\,|\,x_{n+1}=y)\,f_{n+1}(y)dy\;\;P(x_{n+1}\in\Lambda_{n+1}\,|\,x_n=x)\,\text{，}$$

其中 f_{n+1} **是随机变量** x_{n+1} **的分布密度**. **由于上式右方与随机事件** A **无关**, **可见等式左方应该等于**

$P(B|x_n=x)$. **对于一般的** k , **只需做归纳法.）**.

Markov **性的等价性质** 3

在已知"现在"的条件下，"将来"与"过去"　是条件独立的 (**此即**(9. 17)')

Markov **性的等价性质**　4

对随机序列$\{x_n\}$**及**$\forall m\geq 1,n\geq 0$ **及**, **任意** Borel **集** L **及实数** $x,x_0,\cdots,x_{n-1}$, **都有**

$$P(x_{n+m}\in L\,|\,x_n=x,x_{n-1}=x_{n-1},\cdots,x_0=x_0)=P(x_{n+m}\in L\,|\,x_n=x)\,. \tag{9. 18}$$

（直观证明　设一切条件分布密度都存在.　m=1 **时即为** Markov **性的定义**. **对** m **作归纳法，假定** m **时**(9.18)**正确**, **今证** m+1 **时它也正确**:

$$P(x_{n+m+1}\in L\,|\,x_n=x,x_{n-1}=x_{n-1},\cdots,x_0=x_0)$$

$$=\int\;P(x_{n+m+1}\in\Lambda,x_{n+1}=y\,|\,x_n=x,x_{n-1}=x_{n-1},\cdots,x_0=x_0)\,f_{n+1}(y)dy$$

$$=\int\;P(x_{n+m+1}\in\Lambda\,|\,x_{n+1}=y,x_n=x,x_{n-1}=x_{n-1},\cdots,x_0=x_0)$$

$$f_{n+1}(y\,|\,x_n=x,x_{n-1}=x_{n-1},\cdots,x_0=x_0)dy$$

利用归纳法假设及 Markov **性的定义**, **上式简化为**

$$=\int\;P(x_{n+m+1}\in L\,|\,x_{n+1}=y)\;\;f_{n+1}(y\,|\,x_n=x)dy\,.$$

注意上式右方与$\{x_{n-1}=x_{n-1},\cdots,x_0=x_0\}$无关，可知(9.18)式对$m+1$也成立）.

Markov **性的等价性质５**

对于任意有界 Borel **函数** g **，即任意实数** $x,x_0,\cdots,x_{n-1}$ **，均有**

$$E(g(x_{n+1})|x_0=x_0,\cdots,x_{n-1},x_n=x)=E(g(x_{n+1})|x_n=x)\,. \qquad (9.\ 19)$$

(在 $g(x)=I_\Lambda(x)$ 时, (9. 19)就成为(9. 15). 然而，从(9. 15)推导 (9. 19) 需要测度论的基本知识，本书从略．这里我们粗略地指出其朴素思想， 就是$g=g^+-g^-$，其中$g^+=gI_{\{g>0\}}$，而$g^\pm$们可用有限个示性函数的线性组合来近似)．

注 等价性质 1，2，3，5 **诸款都有相应于等价性质** 4 **的形式，读者可自己列出它们**.

Markov **性的等价性质** 6 **（最一般的形式）**

对于任意实 Borel **集** Λ **，及只由随机序列**$\{x_m\}$**在时刻**n**及其后的信息所决定的随机变量**h_n**， 恒有**

$$P(h_n\in\Lambda|x_n=x,x_{n-1}=x_{n-1},\cdots)=P(h_n\in\Lambda|x_n=x) \qquad (9.\ 20)$$

或

$$E(h_n|x_n=x,x_{n-1}=x_{n-1},\cdots,x_0=x_0))=E(h_n|x_n=x)\,. \qquad (9.\ 21)$$

注 这个等价条件的内涵十分丰富， 其等价性的证明在测度论中非常典型，本书从略.

与离散状态的 Markov **链的转移矩阵类似，在连续状态情形，我们有概率转移核**.

2. 3 **概率转移核**

定义９．１７ 记

$$p^{(n,m)}(x,L)=P(x_m\in L\mid x_n=x),(m\ge n)\,. \qquad (9.\ 22)$$

这是一个条件分布(上式右方就是 $E(I_L(x_m)\mid x_n=x)$).

$$p^{(n,n)}(x,L)=d_L(x)\overset{D}{=}\begin{cases}1 & (x\in L)\\ 0 & (x\notin L)\end{cases}\,.$$

特别地， $p^{(n,n+1)}(x,L)$**称为时刻**n **(到时刻**$n+1$**)的概率转移核，而**$p^{(n,n+k)}(x,L)$**就称为从**n**出发的**k**步概率转移核.**

特别地，若存在$p^{(n,m)}(x,y),(n<m)$**，使**

$$p^{(n,m)}(x,L)=\int_L p^{(n,m)}(x,y)dy\,, \qquad (9.\ 23)$$

那么$p^{(n,m)}(.x,y)$**就是：在**$x_n=x$**的条件下，**x_m**的条件分布密度，称为条件转移密度**.

（我们在本书中如无特别声明，则恒假定条件转移密度存在.）它满足以下性质：

（P.1）$0 \le p^{(n,m)}(x,y) \le 1 \quad (n<m)$，$\int p^{(n,m)}(x,y)dy = 1$.

（P.2）（Chapman-Kolmogorov 方程） 对于$\forall \; l \ge m \ge n$ 有

$$p^{(n,l)}(x,y) = \int p^{(n,m)}(x,z)\, p^{(m,l)}(z,y)dz, \qquad (9.24)$$

有时也简写成乘积核的形式，即记成

$$p^{(n,l)} = p^{(n,m)} * p^{(m,l)}.$$

((P.1)是显然的. (P. 2)的直观含义十分清楚，它就是条件概率的全概率公式).

(P. 2)的证明 由 Markov 性及条件期望的性质

$$E(E(I_L(x_l)\,|\,x_m)\,|\,x_n) = E(E(I_L(x_l)\,|\,x_m, x_n)\,|\,x_n) = E(I_L(x_l)\,|\,x_n).$$

从而

$$\iint_{\Lambda} p^{(n,m)}(x,z)\, p^{(m,l)}(z,y)dzdy = E(\int_{\Lambda} p^{(m,l)}(x_m,y)dy\,|\,x_n = x)$$

$$= E(E(I_L(x_l)\,|\,x_m)\,|\,x_n = x) = E(I_L(x_l)\,|\,x_n = x) = \int_{\Lambda} p^{(n,l)}(x,y)dy.$$

（注 当不存在条件分布密度时，Chapman-Kolmogorov 方程应写为

$$p^{(n,l)}(x,L) = \int p^{(n,m)}(x,dy)\, p^{(m,l)}(y,L)), \qquad (9.25)$$

其中积分是 Stieljies 积分）

2. 4 时齐的连续状态 Markov 链

定义９.１８ 连续状态的 Markov 链称为时齐的，如果其概率转移核$p^{(n,n+1)}(x，L)$与n无关(即 1 步转移概率与出发时刻n无关. 我们把它改记为$p(x,L)$.

在时齐且有条件转移密度的情形，我们简记

$$p(x,y) \overset{def}{=} p^{(0,1)}(x,y) = p^{(n,n+1)}(x,y),\; p^{(m)}(x,y) \overset{\Delta}{=} p(0,x;m,y) = p^{(n,n+m)}(x,y)$$

这时 Chapman-Kolmogorov 方程就简化为

$$p^{(n+m)}(x,y) = \int p^{(m)}(x,z)\, p^{(n)}(z,y)dz. \qquad (9.26)$$

在本书中除非特别声明，所考虑的连续状态 Markov 链总是时齐的.

(注 在不存在条件转移密度时，由 (9. 24) 得到

$$p^{(n,n+m)}(x,L) = \left(p^{(n,n+1)} * p^{(n+1,n+2)} * \cdots * p^{(n+m-1,n+m)}\right)(x,L) = p^{(m)}(x,L)$$

它也不依赖出发时刻n， 它是 Markov 链$\{x_n, n \ge 0\}$的m步转移核，其中

$$p^{(2)}(x,L) = \int p(y,L)\,p(x,dy),$$

$$p^{(m)}(x,L) = \int p(y,L)\,p^{(m-1)}(x,dy) = \int p^{(m-1)}(y,L)\,p(x,dy).)$$

在本书中除非特别声明，所考虑的连续状态 Markov 链总是时齐的.

连续状态的 Markov 链的条件转移密度(或不存在密度时的转移核)，是刻画此链的基本统计特征量，一个 Markov 链的统计性质可由其初始状态 x_0（或其初始分布 $p_0(x)=P(x_0=x)$）以及其条件转移密度(或不存在密度时的转移核)完全确定．这就是下面的定理

定理９.１９（Markov链的有限维分布密度） 若连续状态的时齐Markov链 $\{x_n, n\ge 0\}$ 的初始状态为$x_0=x$，其条件转移密度为 $p(x,y)$，则

(1) 在条件 $x_0=x$ 下 $(x_1,\cdots,x_n)$ 的条件联合分布密度为

$$p(x,x_1)p(x_1,x_2)\cdots p(x_{n-1},x_n).$$

如果x_0具有分布密度$j(x)$，那么，$(x_0,\cdots,x_n)$的联合分布密度为

$$j(x)\ p(x,x_1)p(x_1,x_2)\cdots p(x_{n-1},x_n).$$

(2) 对于有界 n 元 Borel 函数 g 有

$$E[g(x_1,\cdots,x_n)\,|\,x_0=x]=$$

$$\int_{\Lambda_n}\cdots\int_{\Lambda_2}\int_{\Lambda_1}p(x,x_1)p(x_1,x_2)\cdots p(x_{n-1},x_n)g(x_1,\cdots,x_n)dx_1dx_2\cdots dx_n.$$

（证明提示 (1)．对于n用归纳法．(2) 对于 $g=g_1(x_1)\cdots g_n(x_n)$ 可用归纳法证明．而对于一般的g，则需要用测度论中的典型逼近方法，这里从略． (其粗略思想为：用示性函数的线性组合近似 $g^+=gI_{\{g>0\}}$.，再用形如$I_{\Lambda_1}(x_1)\cdots I_{\Lambda_n}(x_n)$的函数的线形组合，去近似一般类型的示性函数 $I_\Lambda(x_1,\cdots,x_n)$，其中$L$ 是任意n维 Borel 集)．

注 在转移核不存在密度且x_0没有分布密度的情形，$(x_0,\cdots,x_n)$的联合分布为

$$P(x_0\in\Lambda_0,\cdots,x_n\in\Lambda_n)=\int\limits_{\Lambda_0}\cdots\int\limits_{\Lambda_n}F_0(dx)\ p(x,dx_1)p(x_1,dx_2)\cdots p(x_{n-1},dx_n),$$

其中$F_0(x)$是x_0的分布函数，$p(x,\Lambda)$是 Markov 链的转移核．而

$$E[g(x_1,\cdots,x_n)\,|\,x_0=x]=$$

$$\int_{\Lambda_n}\cdots\int_{\Lambda_2}\int_{\Lambda_1}p(x,dx_1)p(x_1,dx_2)\cdots p(x_{n-1},dx_n)g(x_1,\cdots,x_n).$$

定理９.２０（连续状态的 Markov 链的绝对概率）

设 $f_n(y\,|\,x)$是在条件$x_0=x$下，随机变量x_n的条件分布密度．那么

$$f_{n+m}(y\mid x)=\int f^{(m)}(y\mid z)p^{(n)}(x,z)dz\,. \qquad (9.\ 27)$$

证明　应用全概率公式.

2．5　例

例９．２１　(非随机的迭代映射)　设随机变量序列x_n满足迭代关系

$$x_{n+1}=f(x_n),x_0=x\,.$$

这时x_n没有分布密度．事实上它是一个常数，即由x只能转移到$f(x)$．于是，对于任意实 Borel 集L 可以直接计算得到

$$P(x_{n+1}\in L\mid x_n=x,x_{n-1}=x_{n-1},\cdots,x_0=x_0)$$

$$=P(f(x)\in L\mid x_n=x,x_{n-1}=x_{n-1},\cdots,x_0=x_0)=I_{\{f(x)\in L\}}\,.$$

也就是其转移核为

$$p(x,L)=I_{f^{-1}(L)}(x)\,,$$

其中　$f^{-1}(L)$是L 在f 下的原象集，即$f^{-1}(L)=\{y:f(y)\in L\}$．

例９．２２　(带随机噪声的迭代映射所生成的连续状态的 Markov 链)

设$\{w_n\}$为独立序列，且w_n的分布密度为$f_n(y)$，$g(x,y)$是R^2上的 Borel 函数．又设随机变量序列x_n满足迭代关系

$$x_{n+1}=g(x_n,w_n)\,,$$

其中x_0与$\{w_n\}$独立．　此时w_n，$g(x,w_n)$都与$\{x_0,x_1,\cdots,x_n\}$独立．故而

$$P(x_{n+1}\le y\mid x_n=x,x_{n-1}=x_{n-1},\cdots,x_0=x_0)$$

$$=P(g(x,w_n)\le y\mid x_n=x,x_{n-1}=x_{n-1},\cdots,x_0=x_0)$$

$$=P(g(x,w_n)\le y)=\int_{g(x,z)\le y}f_n(z)dz\,.$$

上式右方与$x_{n-1},\cdots,x_0$无关，从而等于　$P(x_{n+1}\le y\mid x_n=x)$．可见$x_n$是连续状态的Markov 链．而且，当且只当$f_n(y)$不依赖$n$时,这个连续状态的 Markov 链才是时齐的．

2．6　寻找R^d上可微函数$f(x)$的最小值的位置的模拟退火算法

先考虑$d=1$情形．　设w_n为独立同分布的$N(0,1)$随机变量序列，又设函数$f(x)\ge 0$

且二阶光滑，再设 $f(x)$ 取最小值的点全体组成集合 S_0．我们考虑如下的随机迭代模型

$$V_{n+1} = V_n + hf'(V_n) + e_n w_n f(V_n) \qquad (h>0), V_0 = x_0 .$$

它是例９．２２的特殊情形，而且在 $V_n = x$ 的条件下，转移到 $V_{n+1} = y$ 的条件转移密度为

$$p^{(n,n+1)}(x,y) = \frac{1}{e_n \sqrt{2p}\, f(x)} e^{-\frac{1}{2}[\frac{y-x-hf'(x)}{e_n f(x)}]^2} .$$

即它是正态分布 $N(x+hf'(x), e_n{}^2 f(x)^2)$，其中 x 作为参数．于是 V_n 是时间离散的非时齐的状态连续的 Markov 链．这里的非时齐性是由于"小参数" e_n 依赖于时刻 n 而引起的．

对于 $d>1$ 情形，则相应的随机迭代模型为

$$V_{n+1} = V_n + h\nabla f(V_n) + e_n w_n f(V_n) \qquad (h>0), V_0 = x_0 , \qquad (9.28)$$

其中 V_n，w_n 都是随机向量，而 f 为多元函数（注意如果 $V_n \in S_0$，则有 $\nabla f(V_n) = 0$，此时随机迭代模型就变成

$$V_{n+1} = V_n + e_n w_n f(V_n) \qquad (h>0), V_0 = x_0 ,$$

即只有随机小扰动项 $e_n w_n f(V_n)$ 起推动 V_n 运动的作用． 由此可知一旦当此 Markov 链的轨道 V_n 到达了 f 的某个稳定点以后，再到它的给定的某个小邻域以外的可能性就较小，而且当此稳定点是 f 的最小值点时，这种可能性就相对地更小．我们再注意，当此 Markov 链的轨道 V_n 在未到达 f 的某个稳定点之前，在（９．２８）中的项 $h\nabla f(V_n)$ 起的主要作用将使 V_n 往 f 的稳定点发展．综合此两个朴素思想的启示，我们可以用此 Markov 链的发展来"近似"地找到 f 的最小值的位置，这正是连续状态空间变量的模拟退火思想的要点． 当然用这种方法还是有失败的可能的．然而，这样做最终失败的可能性较小．如果我们能忍受小概率失败的失误而放弃，而只求以较大的概率能达到整体最小值附近，那么这是一种可行的妥协方法(这种加入随机小扰动项 $e_n w_n f(V_n)$ 的方法，能避免用确定性方法(例如梯度方法)求最小值位置时,陷入局部极小处的失误)．把 e_n 看成温度后，（９．２８）提供了连续状态空间的一种模拟退火方法，即利用增加小扰动项的随机迭代模型作模拟退火．如果适当地选取退火的温度 e_n 趋于 0 的速度（例如，如 Geman-Gemam 那样选取 $e_n = g(\ln n)^{-1}$，而 g 又足够大），那么，在理论上可以证明：对于任意 $d>0$ 有

$$P(V_n \in B_d(S_0)) \xrightarrow{n\to\infty} 1 ,$$

其中 $B_d(S_0)$ 指 S_0 中所有点的δ邻域的并(这种极限关系称为 V_n 按概率收敛到集合 S_0，与有限状态空间的情形看起来不同,在有限状态空间情形是收敛到最小值集合，而现在时收敛到最小值集合的一个邻域．如果我们认为有限集中的点都是孤立的，那么这两者是一致的)．也就是说，对随机差分方程作迭代，经过足够多次以后，就以近似于 1 的概率，保证 V_n 会到达 f 取最小值的集合的较近的范围．这就是用随机迭代模型作模拟退火的基本思想，这里温度 e_n 依赖 n 的阶 $e_n = g(\ln n)^{-1}$，对于概率收敛性的成立是至关重要的．

注意求函数 $g(x)$ 零点的问题，可以化为求 $g(x)^2$ 的最小值问题，因而也可以应用随机迭代模型的模拟退火方法．因此这个数学模型在广泛的领域中有应用前景．

［注］ 模拟退火方法中的人工随机干扰 w_n，是独立同分布随机序列，称为白噪声．模拟退火的进程一般较慢，对此人们采取了多种加速收敛措施，其中常用的有 TINA (Time Invariant Noise Annealing, 用时间不变的噪声作模拟退火) 算法．它的基本思想是：采用 ″有色噪声″ 作为逃逸出局部极值的 “驱动力″，即在目标函数值大的时候，使干扰大些，而当函数值接近最小值时，就把干扰取得小些．特别是在已知最小值的情形（例如求解非线性联立方程组时，只要把目标函数取为各方程的平方和，那么目标函数的最小值就是 0，而达到最小值的变量就是该方程组的解)．例如，在已知最小值的条件下,对 $f(x)$ 求整体最小值的位置，也可以考虑如下随机差分方程：

$$x_{n+1} = x_n + \nabla f(x_n) + (f(x_n) - f_*)w_n, \qquad (9.29)$$

其中 w_n 是独立同分布的随机变量列， f_* 是目标函数 $f(x)$ 的最小值．于是(９．２９）给出了用有色噪声 $(f(x_n) - f_*)w_n$ 干扰的 TINA 算法．

我们还可以将 TINA 算法修改为适合于未知最小值的全局优化问题，例如可以把干扰噪声取为 $[g(f(x_n) - c)^+) + (g \cdot \ln n)^{-1}]w_n$，其中 g 是一个有界增函数，c 是 一个常数，它取得越接近最小值 f_* 越好．而在实际计算中，人们可以根据对 f 的函数值的认识，不断地调小参数 c，以求加速收敛．

2．7　Dobrushin 不等式.指数遍历性与收敛性

在第 5 章中，我们曾对离散状态的 Markov 链，讨论了当 n →∞时它的极限性态．在本节中，我们将利用 Dobrushin 不等式，对于连续状态的 Markov 链得到相应的结果．这类结果要求假定转移密度存在，并满足较强的条件，这时不仅可以得到 n 步转移密度当 $n\to\infty$ 时的极限趋于唯一的不变概率分布密度，而且保证误差具有指数型的下降速度(称之为指数遍历性)．这一结论比之于一般 Markov 链的遍历定理要强得多．

命题 9．２３　(Dobrushin 不等式)

对于 Borel 函数 f，定义 $\| f \| = \int | f(x) | dx$．假定时间离散的连续状态的 Markov 链

存在转移密度 $p(x,y)$．那么对于任意分布密度 $r_1(x), r_2(x)$ 有

$$\| \int p(x,*) r_1(x)dx - \int p(x,*) r_2(x)dx \| \le C(P) \| r_1 - r_2 \| ， \quad (9.30)$$

其中

$$C(P) = \frac{1}{2}\sup_{x,y} \int |p(x,z) - p(y,z)| dz \quad （9．31）$$

称为转移密度 $p(x,y)$ 的 Dobrushin 数．　　】

例９．２４　设 $f(x)$ 是一个有界连续函数．那么对于满足 $f(x) \ne f(y)$ 的点 (x,y)，

$h(x,y) \overset{L}{=} \frac{f(x)+f(y)}{2}$ 是方程

$$e^{-\frac{(z-f(x))^2}{2}} = e^{-\frac{(z-f(y))^2}{2}}$$

的唯一解，且 $h(x,y)$ 也是有界连续函数．当 $f(x) < f(y)$ 时我们有

$$\int |p(x,z) - p(y,z)| dz \le \int_{-\infty}^{h(x,y)} \frac{1}{\sqrt{2p}} e^{-\frac{(z-f(x))^2}{2}} dz + \int_{h(x,y)}^{\infty} \frac{1}{\sqrt{2p}} e^{-\frac{(z-f(y))^2}{2}} dz$$

$$= \int_{-\infty}^{h(x,y)-f(x)} \frac{1}{\sqrt{2p}} e^{-\frac{u^2}{2}} du + \int_{h(x,y)-f(y)}^{\infty} \frac{1}{\sqrt{2p}} e^{-\frac{u^2}{2}} du$$

$$\le \int_{-\infty}^{m_2} \frac{1}{\sqrt{2p}} e^{-\frac{u^2}{2}} du + \int_{m_1}^{\infty} \frac{1}{\sqrt{2p}} e^{-\frac{u^2}{2}} du = \Phi(m_2) + (1 - \Phi(m_1))，$$

其中

$$m_1 \overset{D}{=} \inf_{x,y} \{|h(x,y) - f(x)| : x \ne y\}，\ m_2 \overset{D}{=} \sup_{x,y} \{|h(x,y) - f(x)| : x \ne y\}，$$

$$F(a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2p}} e^{-\frac{u^2}{2}} du．$$

显见 $-\infty < m_1 < m_2 < \infty$，于是

$$C(P) = \frac{1}{2}\sup_{x,y} \frac{1}{\sqrt{2p}} \int \left| e^{-\frac{(z-f(x))^2}{2}} - e^{-\frac{(z-f(y))^2}{2}} \right| dz$$

$$\le \frac{1}{2}[\Phi(m_2) + (1 - \Phi(m_1))] < 1．$$

例９．２５　设连续状态的 Markov 链的转移密度 $p(x,y)$ 满足条件

$$p(x,y) \ge g(y) \ge 0，$$

那么此转移密度的 Dobrushin 数满足

$$C(P) \le 1 - \int g(y)dy \le 1.$$

证明　仿照第５章离散状态的 Markov 链的 Dobrushin 不等式的证明可知

$$C(P) = \frac{1}{2}\sup_{x,y}\int |p(x,z) - p(y,z)|dz =$$

$$= \sup_{x,y}\int_{p(x,z)\ge p(y,z)} (p(x,z) - p(y,z))dz \le \sup_{x,y}\int_{p(x,z)\ge p(y,z)} (p(x,z) - g(z))dz$$

$$\le \sup_{x,y}\int (p(x,z) - g(z))dz = 1 - \int g(z)dz.$$

定义９．２６　具有转移密度$p(x,y)$的连续状态的时齐 Markov 链x_n 称为指数 L^1遍历的(或强指数遍历的)，如果存在正整数n_0，分布密度 $r^{(\infty)}(y)$ 及常数$0<r<1, c>0$，只要$n \ge n_0$，就对任意x满足

$$\int |p^{(n)}(x,y) - r^{(\infty)}(y)|dy < cr^n,$$

其中 $p^{(n)}(x,y)$是此 Markov 链的n步转移密度.

下面的概念只需要转移核，而不需要假定存在转移密度.

定义９．２７　连续状态的时齐马氏链x_n称为指数遍历的，如果存在正整数n_0，分布函数$F^{(\infty)}(y)$及常数$0<r<1, c>0$，只要$n \ge n_0$，就对任意x，y满足

$$|P(x_n(x) \le y) - F^{(\infty)}(y)| < cr^n,$$

这里改用$x_n(x)$记初值为$x_0 = x$的 Markov 链x_n．分布函数 $F^{(\infty)}(y)$不依赖于初值x，称为x_n的(遍历)极限分布，也称为x_n的平稳分布．以$F^{(\infty)}$为初分布，就能得到一个连续状态的平稳过程．显见，指数L^1遍历一定指数遍历.

定理９．２８　(指数 L^1遍历定理)

具有转移密度的连续状态时齐 Markov 链x_n，如果存在$n_0 \ge 1$，使其n_0步转移密度$p^{(n_0)}(x,y)$的 Dobrushin 数$C(P^{(n_0)}) < 1$，那么此 Markov 链是指数 L^1遍历的.

(证明与第 5 章中离散状态情形类似,从略).

例９．２９　例９．２５中的 Markov 链, 只要存在$n_0 \ge 1$, 使其n_0步转移密度满足

$$p^{(n_0)}(x,y) \ge g(y), \text{ 且} \int g(y)dy > 0,$$

则它指数 L^1遍历

例９．３０　设 f 是有界函数，而 w_n 是独立同分布随机变量序列，且具有严格正的连续的分布密度 $p(x)$．那么由如下定义的迭代列 x_n 所决定的连续状态 Markov 链是指数 L^1 遍历的

$$x_{n+1}=f(x_n)+w_n. \qquad (9.32)$$

证明　不妨设 $a\le f(x)\le b$．此 Markov 链的转移密度为

$$p(x,y)=p(y-f(x))\ge \min_{a\le z\le b} p(y-z)>0.$$

由例９.２５和定理９．２８可知此 Markov 链是指数 L^1 遍历的.

定理９.３１（指数遍历定理）

设 w_n 是独立同分布随机变量序列，$f(u,v)$ 是连续函数，且关于 u 单调不减．又

$$x_0=x, x_{n+1}(x)=f(x_n(x),w_n). \qquad (9.33)$$

如果存在 z_0，n_0 及 $d>0$ 使

$$P(\sup_x x_{n_0}(x)\le z_0) \text{ 及 } P(\inf_x x_{n_0}(x)\ge z_0) \text{ 都} \ge d，$$

则连续状态的 Markov 链 $x_n(x)$ 是指数遍历的.

(证略).

3　随机的迭代函数系统（随机 IFS，随机 Iterative Function System）

3．1 局部相似性的基本想法

如第２节所示，图象的局部相似性是它的基本特点．　最简单的图象是只有黑白两个灰度的轮廓图.由于这种图的轮廓线在局部总有一点相似性(所谓自相似性),人们就希望用函数迭代映射的 ”不动点”，来近似描绘轮廓线图．并进一步把一般的黑白灰度图和彩色图用随机迭代函数映射的不变分布来近似.这样就可以达到在传输过程中压缩图象信息,而在接收后再重建合成人工图象的目的.(显然这是非常朴素的想法，例如，压缩图象信息的压缩比如果能达到 20，则效果就很不错了．但是在实践中，还必须考虑到与原有的设备能相容这个方法,即改进压缩比的方法只希望在原有设备上附加一个系统,而不希望完全改建设备．因此再理想的压缩方法也必须与实际情况相协调).

图像的纹理分析(Texture Analysis)是一种重要的图像处理方法, 在目标识别等方面有重要应用。图象的纹理是一种结构, 它包括：自然纹理表面的灰度，颜色，条纹, 以及它们的模拟. 要模拟或分析图像纹理，选择合适的模型是关键的一步. 纹理模型必须能表述图像元素的空间分布特征. 均匀的纹理的最基本的分布特性是它的空间相似性，就是在图像上的不同地方的纹理有着几何上相似的象素组合. 在数学上，可以用小波分析, 自相似随机过程, 随机迭代函数系统, Markov 随机场等来描述纹理. 由于 Markov 随机场和 Gibbs 随机场是等价的，因而就可以用 Gibbs 随机场模型 (用另一种看法视为随机图, 而把图像作为随机图的

一个样本）描述纹理．进而可以使用各种方法估计 Gibbs 分布的参数，在此基础上就可以对纹理进行分析和处理．下面介绍的轮廓图的方法，是一种与 Gibbs 分布不同的，近似描绘无灰度的黑白图的方法，即迭代函数系统方法．

３．２　轮廓图全体组成的距离空间

定义９．３２（轮廓图的图象窗口中的距离）

只有轮廓而无灰度的黑白图，即轮廓图，其黑色部分就可看成为平面上某个窗口的一个闭集．我们记此窗口为 X，它是 R^2 中的有界闭集，例如，$X=[0,1]\times[0,1]$．X 中的两个点 $x=(x_1,x_2), y=(y_1,y_2)$ 之间可定义距离 $d: X\times X\to[0,\infty)$，例如普通的 Euclid 距离 $d(x,y)=\sqrt{(x_1-y_1)^2+(x_2-y_2)^2}$（也可以是其它的距离，只要满足如下的 3 条距离公理：

（D.1）$d(x,y)\geq 0$，而且等号成立的充要条件为 $x=y$；

（D.2）（对称性）$d(x,y)=d(y,x)$；

（D.3）（三角形不等式）$d(x,z)\leq d(x,y)+d(y,z)$．

例如，$d(x,y)=\max(|x_1-y_1|,|x_2-y_2|)$，或者 $d(x,y)=|x_1-y_1|+|x_2-y_2|$）．

轮廓图的图象空间与图像之间的距离

如上所述，一幅轮廓图就是窗口的一个闭子集．从数学上描述两幅轮廓图的近似程度的最朴素的想法是，在两幅图之间引入距离（也可以引入准距离“相对熵”，这将在第 10 章的隐 Markov 模型与第 15 章的算法中叙述）．为此，我们先介绍一些关于距离空间的基本知识．由前面的论述，我们知道，窗口 X 可配以距离 d，我们把 (X,d) 作为蓝本，引入距离空间的概念如下．

定义９．３３　任意一个集合 Y，如果定义了满足 3 条距离公理的映射 $d: Y\times Y\to[0,\infty)$，则我们称 (Y,d) 为距离空间（metric space）．距离空间中有收敛概念，对于 $x_n\in Y$，如果有 $d(x_n,x)\to 0$，则称 x_n 收敛到 $x\in Y$，记为 $x_n\to x$．

距离空间中的序列 $\{x_n\},(x_n\in Y)$ 称为 Cauchy 列，如果对于任给 $e>0$，存在 N，只要 $n,m>N$，就有 $d(x_n,x_m)<e$（显见这个概念是欧氏空间中的 Cauchy 序列的推广．需要注意的是，在距离空间中并没有加法运算，除非它还具有线性空间结构）．

距离空间 (Y,d) 称为完备的，如果它的任意 Cauchy 列 $\{x_n\}$ 都有极限，即存在 $x\in Y$，使 $d(x_n,x)\to 0$．（例如，R^d，R^d 中带边的超立方体，R^d 中带边的有界图形，都是完

备的距离空间).

对于距离空间(Y,d)，Y 的子集F 称为闭集，如果F 中的任意收敛序列的极限仍在 F 之中. 闭集的余集称为开集．Y 的子集K 称为紧集，如果K 中的任意序列都有收敛子序列，而且此收敛子列的极限仍在K 之中.

显见紧集必是闭集．在欧氏空间中，紧集与有界闭集是一样的.在上述的窗口空间中，由于窗口本身是有界的，因而闭集与紧集是一样的．但是，在一般距离空间闭集就未必是紧集.

定义９.３４　（作为距离空间的轮廓图的图象空间）

设(X,d)为窗口距离空间．在此窗口空间上的一个轮廓图，就是窗口空间的一个闭集．记

$$H(X)\overset{\Delta}{=}\{F:F\text{闭集}\subset X\}\text{.}$$

那么．$H(X)$为所有的轮廓图的图象组成的集合，称为图象空间．对于任意两个轮廓图 $A,B\in H(X)$，可以定义如下三种不同的距离 r, r_1, r_2：

$$r(A,B)\overset{\Delta}{=}\sup_{x\in A} d(x,B)+\sup_{y\in B} d(y,A),$$

其中

$$d(x,B)=\inf_{y\in B} d(x,y);$$

$$r_1(A,B)\overset{\Delta}{=}\inf_d\{d: A\subset B^d, B\subset A^d\},$$

其中

$$A^d=\{y:\exists x\in A\text{使}d(x,y)<d\};$$

$$r_2(A,B)\overset{\Delta}{=}[\sup_{x\in A} d(x,B)]\vee[\sup_{y\in B} d(y,B)], \qquad (9.\ 34)$$

其中

$$a\vee b=\max(a,b).$$

不难验证 r, r_1, r_2 都是$H(X)$上的距离，而且它们都彼此等价，即收敛性是一样的(虽然距离的大小并不一样).

命题９．２５　距离空间$(H(X),r)$是完备的.

证明　设A_n为$(H(X),r)$的 Cauchy 列，则易见对于任意$x_n\in A_n$，x_n是(X,d)的 Cauchy 列，故有极限x．令

$$A\overset{\Delta}{=}\{x:\exists x_n(\in A_n)\to x\}.$$

用与普通数学分析类似的方法，可以证明 A是闭集，即$A\in H(X)$．而且 $r(A_n,A)\to 0$．这就证明了

$(H(X), r)$ 的完备性.

完备距离空间的压缩映射与不动点

定义９．２６　任意距离空间 (X,d) **自身的映射** W，**如果存在** $c<1$，**使得对于任意** $x, y \in X$，**恒有**

$$d(W(x),W(y)) \le cd(x,y). \tag{9.35}$$

那么，W **称为压缩映射**，c **称为压缩系数**，

例９．２７（仿射压缩映射）设 W **为如下映射**

$$x=\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \xrightarrow{W} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}+\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}.$$

假定系数 $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$，$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ **满足**：$\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ **的** W **像都在** X **中**，**且** $c \overset{\Delta}{=} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}^T$ **的最大特征值** <1 **．那么，可以证明这样的仿射映射是压缩映射**.

定理９．２８（不动点定理）　对于任意完备距离空间 (X,d) **上的压缩映射** W **有**

(1)　**存在唯一的不动点** $a: W(a)=a$.

(2)　(**迭代近似**) **记** $W^{\circ 2}=W\circ W$，$W^{\circ 2}(x)=W(W(x)), W^{\circ n+1}=W\circ W^{\circ n}$，**则对于任意** $x\in X$，**恒有** $W^{\circ n}(x)\to a$.

(3)　(**近似不动点**) **如果** y **满足** $d(W(y),y)<e$，**则** $d(y,a)\le \dfrac{e}{1-c}$.

证明　先证明不动点的存在性．令 $x_n=W^{\circ n}(x)$，**于是** $d(x_{n+1},x_n)\le c^n d(x_1,x)$，**从而**

$$d(x_{n+p},x_n)\le \sum_{k=1}^{p} d(x_{n+k},x_{n+k-1})\le \frac{c^n}{1-c}d(x_1,x)\to 0.$$

所以 $\{x_n\}$ **是** Cauchy **列**. **由完备性**，**它有极限**，**记为** a. **再则**

$$d(W(a),a)\le d(W(a),W(x_n))+d(W(x_n),x_n)+d(x_n,a)\to 0,$$

故 $W(a)=a$. **再证明不动点只有一个．假设另有不动点** b，**则有**

$$d(a,b)=d(W(a),W(b))\le cd(a,b).$$

由$c<1$立得 $d(a,b)=0$，即 $a=b$．这就证明了(1)和(2)．最后

$$d(y,a) \le d(y,W(y)) + \cdots + d(W^{\circ n-1}(y), W^{\circ n}(y)) + d(W^{\circ n}(y), a)$$

$$\le e + ce + \cdots + c^{n-1}e + d(W^{\circ n}(y), a) \le \frac{e}{1-c}.$$ 】

从窗口空间(X,d)的压缩映射导出图象空间$(H(X),r_1)$的压缩映射

定理２．２９（Collage 拼贴定理）

若窗口空间(X,d)上有有限个压缩系数为c_i的压缩映射W_i $(1 \le i \le N)$．定义$(H(X),r_1)$上的映射$\hat{W}$，其象集为所有的$W_i(i \le N)$的象集之并：

$$\forall A \in H(X)，\hat{W}(A) \overset{\Delta}{=} \bigcup_{i=1}^{N} W_i(A).$$

那么，$\hat{W}$是$(H(X),r)$上的压缩映射．$\hat{W}$称为由$\{W_1,\cdots,W_N\}$导出的迭代函数系统．它的唯一不动点A_0称为迭代函数系统$\hat{W}$的吸引子．而且对于任意$F \in H(X)$有误差估计(称为拼贴定理)（设c_i们中的最大数为c）

$$r_1(F,A_0) \le \frac{1}{1-c} r_1(F,\hat{W}(F)). \qquad (9.36)$$

证明　由$H(X)$上的等价距离r_1的定义得到

$$r_1(\hat{W}(A),\hat{W}(B)) = r_1(\bigcup_{i=1}^{N} W_i(A), \bigcup_{i=1}^{N} W_i(B))$$

$$\le \max_i r_1(W_i(A),W_i(B)) \le c r_1(A,B).$$ 】

迭代函数系统的吸引子的"自相似"特性

由$\{W_1,\cdots,W_N\}$导出的迭代函数系统$\hat{W}$的吸引子A_0满足：

$$W_i(A_0) \subset \bigcup_{j=1}^{N} W_j(A_0) = \hat{W}(A_0) = A_0.$$

这说明吸引子A_0的W_i象是A_0的一部分．再用W_k们继续地迭置作用下去，则它相当于对A_0不断地作细分．在国内外工程文献中，称以上性质为：迭代函数系统$\hat{W}$的吸引子A_0关于$\{W_1,\cdots,W_N\}$有自相似性．事实上，当W_k们都是仿射压缩映射时，这种说法较为合理，而在一般情形是一种近似的描述．　在实际运作中，迭代函数系统中的映射常取成仿射压缩

映射，这就充分反映了轮廓图的自相似性．

用迭代函数系统压缩图象信息的思路为：通过选取 N 及 $\{W_1,\cdots,W_N\}$，用迭代函数系统 $\hat{W}$ 的吸引子(在距离 r_1 下)去近似标的图象．由此可以达到压缩信息的目的，因为这时轮廓图的信息完全可由 $\{W_1,\cdots,W_N\}$ 描述．下面的定理说明上述要求不但能够做到，而且可以把 W_i 们选取为仿射映射．

定理９．３０ (Hutchinson 定理)

任给的轮廓图 $F\in H(X)$ 及容许误差 $e>0$，存在 $m>1$，及压缩的仿射映射组 $\{W_1,\cdots,W_m\}$，使 $\hat{W}$ 的吸引子 A_0 是 F 的容许近似，即

$$r_1(F,A_0)\le e\ .$$

证明　因为 F 是紧集，所以可以用有限个半径为 $\frac{e}{8}$ 的圆 $B_1,\cdots,B_m$ 覆盖 F：$F\subset\bigcup_{k=1}^{m}B_k$．我们不妨假定这些圆的中心都在 F 的 $\frac{e}{8}$ 邻域中，于是 $F\subset\bigcup_{k=1}^{m}B_k\subset F^{\frac{e}{4}}$．我们可以找压缩系数 $c_i<\frac{1}{2}$ 的仿射映射 W_i，使 $W_i(F)\subset B_i$．又由于 B_i 的直径小于 $\frac{e}{4}$，故

$$W_i(F)\subset B_i\subset W_i(F)^{\frac{e}{2}}.$$

从而

$$\hat{W}(F)\subset F^{\frac{e}{4}},\ \ F\subset\hat{W}(F)^{\frac{e}{2}}.$$

因此，由距离 r_1 的定义得到

$$r_1(F,\hat{W}(F))\le\frac{e}{2}$$

再用 Collage 定理推出

$$r_1(F,A_0)\le e\,.$$

[注]　Hutchimson 定理出现于 1981 年. 它的主要含义是: 可以把轮廓图近似地作为迭代函数系统的吸引子, 吸引子具有 “自相似性”, 也就是工程界广泛认为的 ”分形性质”.

轮廓图是窗口 X 的一个闭子集，可以用迭代函数系统的吸引子来建模．而窗口上的灰度图的相对灰度可以看成口上的一个概率分布，为此考虑用随机迭代函数系统生成的 Markov 链的不变分布来建模．这就是下一段的理论.

在窗口 $X=[0,1]\times[0,1]$, 相对灰度图就相应地考虑为概率空间 $(X,\mathrm{B}(X))$ 上的一个概率分布，其中 $\mathrm{B}(X)$ 为 X 的 Borel 子集全体组成的类．

3.3　**灰度图与随机迭代函数系统**

本段考虑最简单的随机映射及由它生成的 Markov **链，并用以建模．这种随机映射就是：对于有限个映射，以不同的概率从中取一个．其确切的定义为**

定义９．３１　**窗口** $X=[0,1]\times[0,1]$ **自身的随机压缩映射 $\mathbf{W}$ 的定义为**: **存在 X 自身的压缩映射** $W^{(1)},\cdots,W^{(m)}$，**使**

$$P(\mathbf{W}=W^{(i)})=p_i,\quad (p_1+\cdots+p_m=1).$$

直观地，我们可以将它记为如下的分布列的形式（与离散分布的不同处是其取值为映射，而不是数）

$$\mathbf{W}\sim\begin{pmatrix} p_1,\cdots,p_m \\ W^{(1)},\cdots,W^{(m)} \end{pmatrix}. \tag{9.37}$$

我们称 $\begin{pmatrix} p_1,\cdots,p_m \\ W^{(1)},\cdots,W^{(m)} \end{pmatrix}$ **为随机映射 $\mathbf{W}$ 的分布**.

命题９.３２（独立同分布的随机映射生成的 Markov **链）**

设 $\mathbf{W}_1,\cdots,\mathbf{W}_n,\cdots$ **为独立同分布的随机映射序列**. **设** x_0 **是随机变量．定义**

$$x_1=\mathbf{W}_1(x_0),\cdots,x_{n+1}=\mathbf{W}_{n+1}(x_n), \tag{9.38}$$

则 $\{x_n\}$ **是时齐的** Markov **链**, **称为随机迭代函数系统，其概率转移核(不存在条件转移密度)为**

$$p(x,L)=\sum_{i=1}^{m}p_i I_{(W^{(i)})^{-1}L}(x). \tag{9.39}$$

证明

$$P(x_{n+1}\in L\,|\,x_n=x,x_{n-1}=x_{n-1},\cdots,x_0=x_0)$$

$$=P(\mathbf{W}_{n+1}(x)\in L\,|\,x_n=x,x_{n-1}=x_{n-1},\cdots,x_0=x_0)$$

$$=P(\mathbf{W}_{n+1}(x)\in L)=\sum_{i=1}^{m}p_iP(W^{(i)}(x)\in L)=\sum_{i=1}^{m}p_iI_{(W^{(i)})^{-1}L}(x).$$

推论９．３３

$$E[f(x_{n+1})\,|\,x_n=x]=\sum_{i=1}^{m}p_i f(W^{(i)}(x)).$$ 】

下面介绍独立同分布的随机压缩映射生成的 Markov **链不变分布的存在与唯一性．**

定理９．３４（遍历性定理）

随机压缩映射生成的 Markov **链** x_n **存在唯一的不变分布**. **设它是某个随机变量** V **的分**

布函数 $F_V(x)$， 那么对于此 Markov 链的转移核 $p(x,L)$，分布函数 F_V 有以下的不变性质:

$$F_V(x)=\int p(y,(-\infty,x])dF_V(y),$$

其中 $x=(x_1,x_2),(-\infty,x]\overset{\Delta}{=}(-\infty,x_1]\times(-\infty,x_2]$，$F_V$ 称为不变分布函数． 而且，对于任意有界 Borel 函数 f，不管什么初值 x_0，恒有遍历定理，即有如下的按概率收敛式

$$\frac{1}{n}\sum_{i=1}^{n} f(x_k)\xrightarrow{p} Ef(V)\overset{\Delta}{=}\int f(x)dF_V(x) \qquad (n\to\infty).$$

特别地，如果把 $\{x_k\}_{1\le k\le n}$ 中落入 $(-\infty,x]$ 的频数记为 $n_{(-\infty,n]}$，则有以下的按概率收敛式：

$$\frac{n_{(-\infty,x]}}{n}\xrightarrow{p} F_V(x) \qquad (n\to\infty),$$

这就得到了不变分布函数的相合估计如下(如果将一幅灰度图用一个随机迭代函数系统建模，那么，利用此相合估计可以在实际计算中，通过随机迭代函数系统重建灰度图)

$$\hat{F}_V(x)=\frac{n_{(-\infty,x]}}{n}. \qquad (9.\ 40)$$

证明　令 $h_0=x_0=x$． 定义

$$h_1=\mathbf{W}_{\mathbf{n}+1}(h_0),h_2=\mathbf{W}_{\mathbf{n}}(h_1),\cdots,h_{n+1}=\mathbf{W}_1(h_{\mathbf{n}}).$$

那么 $\{h_n\}$ 与 $\{x_n\}$ 同分布．由于 $\{h_n\}$ 是某个概率空间 (Ω,F,P) 上的随机变量序列，对固定某个基本事件(样本) $w\in\Omega$，存在 $1\le i_1,\cdots,i_n\le m$ 使 $h_n(w)=W^{(i_1)}\circ\cdots\circ W^{(i_n)}(x)$ (其中运算 $\circ$ 表示映射的复合)．于是对于此固定的基本事件 w，压缩映射 $W^{(1)},\cdots,W^{(m)}$ 中最大的压缩系数 $c(<1)$ 有

$$d(h_{n+p}(w),h_n(w))=d(W^{(i_1)}\circ\cdots\circ W^{(i_n)}(x),(W^{(i_1)}\circ\cdots\circ W^{(i_{n+p})}(x))$$

$$\le c^n d(x,W^{(i_{n+1})}\circ\cdots\circ W^{(i_{n+p})}(x))\to 0.$$

这说明 $\{h_n(w))$ 是 X 中的 Cauchy 列，从而它收敛于某个(依赖于初值 x 的) $V^x(w)$，即存在随机变量 V^x (理论上可以严格证明 $\{V^x\le a\}\in$ F)，使

$$P(h_n\xrightarrow{n\to\infty}V^x)=1.$$

再则，由于 $h_n(w)$ 也依赖于初值 x，我们将它改记为 $h_n^x(w)$．又因为它是压缩映射的复合， 所以仍然是压缩映射，故有

$$d(h_n^x(w),h_n^y(w)) \le c^n d(x,y) \to 0 \quad \text{(关于}w\text{是一致的)},$$

由此得到

$$d(V^x(w),V^y(w)) \le d(V^x(w),h_n^x(w)) + d(h_n^x(w),h_n^y(w)) + d(h_n^y(w),V^y(w)) \to 0.$$

这说明极限随机变量V^x与初值x无关，所以我们可以将它改记为V. 显见，h_n的分布函数F_{h_n}按分布收敛到V的分布函数F_V（即在$F_V(x)$的连续点x上有$F_{h_n}(x) \to F_V(x)$).

由x_n与h_n同分布得到， x_n的分布函数F_{x_n}也按分布收敛到V的分布函数F_V. 可以证明: 这个分布F_V就是转移核$p(x,L)$的唯一的不变分布函数.

(注 此事实可以证明如下: 对于任意有界连续函数f，有 $f(x_{n+1}) = f(\mathbf{W}_{n+1}(x_n))$)

$$Ef(\mathbf{W}_{n+1}(x_n)) = \sum_{i=1}^m p_i Ef(W^{(i)}(x_n)) = E[\sum_{i=1}^m p_i f(W^{(i)}(x_n))]$$

$$= \int \sum_{i=1}^m p_i f(W^{(i)}(x))dF_{x_n}(x) = \int \sum_{i=1}^m p_i f(x)dF_{x_n}((W^{(i)})^{-1}x).$$

但是按定义有$f(x_{n+1}) = f(\mathbf{W}_{n+1}(x_n))$, 对此两边取期望便得到

$$\int f(x)dF_{x_{n+1}}(x) = \int f(x)d\sum_{i=1}^m p_i F_{x_n}((W^{(i)})^{-1}x).$$

令$n \to \infty$, 利用按分布收敛的性质推出

$$\int f(x)dF_V(x) = \int f(x)d\sum_{i=1}^m p_i F_V((W^{(i)})^{-1}x).$$

由f的任意性， 再利用（9.39）式便得到

$$F_V(x) = \sum_{i=1}^m p_i F_V((W^{(i)})^{-1}x) = \sum_{i=1}^m p_i \int I_{(W^{(i)})^{-1}(-\infty,x]}(y)dF_V(y) = \int p(y,(-\infty,x])dF_V(y).$$

(把它写成测度的形式就是

$$F_V(L) = \int p(y,L)dF_V(y)).$$

这就证明了极限随机变量V的分布F_V是转移核$p(x,L)$的不变分布函数. 再则, 如果此转移核另有不变分布函数F，将以它为分布函数的随机变量记为J，取$x_0 = J$，那么, 由迭代定义的随机变量序列$\{x_n\}$就是一个平稳序列. 与上面的证明类似地仍有$P(x_n \to V) = 1$. 于是作为极限分布的F应该等于

F_V. 此外，还是由于 $P(x_n \to V)=1$ 及平稳列的性质，还可以进一步证明有不依赖初值的遍历定理). 】

(对于具有一些抽象积分知识的读者，我们还可以用距离空间的方法来证明不变分布函数的存在性如下:

令 M 为窗口空间 (X,d) 上全体概率测度的集合，即 M 中的每一个元素 m 都是 X 上的 Borel 集合类 B(X) 上定义的一个概率. 在 M 上可以定义如下的 Hutchinson 距离 r：

$$r(m,n) \overset{\Delta}{=} \sup_{f\in L} \left| \int f(x)dm(x) - \int f(x)dn(x) \right|,$$

其中 $L \overset{\Delta}{=} \{f : |f(x)-f(y)| \le d(x,y), (\forall x,y \in X)\}$. 可以证明(M，$r$)是完备距离空间. 对于压缩系数为 c_i 的压缩映射 $W^{(i)}$ 们组成的随机映射 $\begin{pmatrix} p_1,\cdots,p_m \\ W^{(1)},\cdots,W^{(m)} \end{pmatrix}$，我们定义它在 M 上导出的一个映射为 (记为 W^{-1})

$$m \to W^{-1}m \overset{\Delta}{=} \sum_{i=1}^{m} p_i (W^{(i)})^{-1} m.$$

往证 W^{-1} 是压缩的.

$$r(W^{-1}m, W^{-1}n) = \sup_{f\in L} \left| \int f(x)d(W^{-1}m)(x) - \int f(x)d(W^{-1}n)(x) \right|$$

$$= \sup_{f\in L} \left| \sum_{i=1}^{m} p_i \int f(W^{(i)}(x))dm(x) - \int f(W^{(i)}(x))dn(x) \right|$$

$$\le \sum_{i=1}^{m} c_i p_i \sup_{f\in L} \left| \int \frac{1}{c_i} f(W^{(i)}(x))dm(x) - \int \frac{1}{c_i} f(W^{(i)}(x))dn(x) \right|$$

$$\le \sum_{i=1}^{m} c_i p_i r(m,n) \le c\, r(m,n),$$

这里我们用了：对于 $f \in L$，有 $\frac{1}{c_i} f \circ W^{(i)} \in L$，而 c 为 c_i 们中的最大者. 一旦证明了压缩性，那么 W^{-1} 作为完备距离空间(M，r)的迭代映射的不动点就存在唯一，而且可以用迭代逼近，这个不动点就是我们所要的不变分布函数).

遍历定理中关于压缩映射的条件可以减弱为如下的平均压缩性. 这就是 Barnsley 与 Elton 在 1988 年得出的平均压缩条件下的遍历定理

定理９．３５ 若随机映射 $\begin{pmatrix} p_1,\cdots,p_m \\ W^{(1)},\cdots,W^{(m)} \end{pmatrix}$ 满足平均压缩条件: 存在 $c<1$，使对于任

意 $x, y \in X$ 有

$$\prod_{i=1}^{m}[d(W_i(x),W_i(y))]^{p_i} \le cd(x,y), \tag{9.41}$$

那么，对于由独立同分布的如上的随机映射生成的随机的迭代函数系统，遍历定理的所有结论仍然成立.

(证明略).

(以上的不动点(不变分布函数)，作为图像有某些自相似性. Barnsley 等人进一步希望用某种不动点来描述自相似的图象，为此提出了递归迭代函数系统(Recursive Iterative Function System)，他用一个 Markov 转移矩阵 P 来代替随机迭代函数系统中的随机向量 $(p_1,\cdots,p_m)$. 即记以 P 为转移矩阵的时齐的 Markov 链为 $\{h_n\}$，再定义

$$x_{n+1} = W^{(h_n)}(x_n).$$

此时 $\{x_n\}$ 不再是 Markov 链，但是 $\{x_n, h_n\}$ 却是 Markov 链. 同样可以得到它的不变分布函数.

另一方面，我们还可以在窗口上的全体概率测度的集合上，定义一个压缩映射得到不动点，而且它与上面定义的是一样的. 进一步还可以考虑有多个递归迭代函数的系统，得到这个系统的不动点).

4 统计中的 Bayes 方法与图象的处理，分割与重建

4.1 Bayes 统计要义

Bayes 方法的基本想法是：把概率函数（包括分布密度和离散的分布概率）中的未知参数 q 当作随机变量（或随机向量).

在 Bayes 方法中不再区分参数与随机变量．所以未知参数的分布的确认是最重要的．在抽取样本以前，就只能根据先验知识设置未知参数的分布，称为先验分布，或验前分布．在抽取样本以后，根据对于抽取到的样本的概率规律的了解，就可用 Beyes 公式把参数先验分布改进为后验分布，或验后分布，也称 Bayes 分布．

1．先验分布与后验分布（Bayes 分布)

设随机变量 x（或随机向量）的分布（或概率函数）为 $p(x,q)$（q 可以是向量)．用 Bayes 的观点，把 q 看成随机变量后，$p(x,q)$ 就是：在 q 已知的条件下，x 的条件分布，即

$$p_{x|q}(x\,|\,q) = p(x,q)\,.$$

假定 q 的分布密度（或概率函数）为 $f(J)$，称为 q 的先验分布（**a priori distribution**),.

于是 (x,q) 的联合分布为 $f(q)\,p(x,q)$．在 J 取固定值的条件下，假若 $x_1,\cdots,x_n$ 为采自 $p(x,q)$ 的独立随机样本． 在 $x_1,\cdots,x_n$ 已知的条件下，q 的条件分布记为

$j(J \mid x_1,\cdots,x_n)$．由 Bayes 公式可知它是q 与$(x_1,\cdots,x_n)$的联合密度$f(q)\prod_{i=1}^{n} p(x_i,q)$，除以$(x_1,\cdots,x_n)$的边缘密度，即

$$j(J \mid x_1,\cdots,x_n)\ \frac{f(J)\prod_{i=1}^{n} p(x_i,J)}{\int f(J)\prod_{i=1}^{n} p(x_i,J)dJ}, \tag{9. 42}$$

称为q 的后验分布（ **a posteriori distribution**），或 Bayes 分布.

2． **已知损失函数是的的 Bayes 估计**

假定对参数J 的一个估计$\hat{J}$ 带来的损失可由一个损失函数$L(J,\hat{J})$来度量．我们用E_{Bayes}表示在后验分布下取的数学期望． 对于J 的任意一个估计量$\hat{J}$， 它在后验分布下的平均损失$E_{Bayes}L(J,\hat{J})$称为后验风险．使后验风险达到最小的估计, 称为 Bayes 估计，记为$\hat{J}_{Bayes}$。即它满足

$$E_{Bayes}[L(J,\hat{J}_{Bayes})] = \min_{\hat{J}} E_{Bayes}[L(J,\hat{J})]\,. \tag{9．43}$$

3． **Bayes 方法评述**

Bayes 方法的长处是设置了一个先验分布，它可以发挥对未知参数的已有统计知识的认知．然而其难点也正在于如何选取合适的先验分布. 这可以说是这种由从另一个角度认识统计带来的一对矛盾. 在实际应用中正是要巧妙地处理好这一对矛盾．

再则，由于在 Bayes 方法中，参数与随机变量是平等的，所以对于随机变量的估计也常用 Bayes 估计．特别是在数据不全的情形（缺失数据）的估计中，需要用测量数据来估计缺失数据，或者是在滤波问题中，或隐 Markov 模型中，需要用量测数据估计状态 （参见第 10 章与第 15 章）．

4．**先验分布的参考取法**

第一种取法出现在无任何先验信息的情形．具体地说

（1）在J 取值于$(-\infty,\infty)$时，认为它是均匀的（这称为 Bayes 假设），即先验密度为常数 (如果q 可能取值区域面积为无限, 则它不是真正的分布, 而是所谓的"广义"分布，但是此时(9. 43)仍然是一个分布，仍称为后验分布或 Bayes 分布), 那么, Bayes 估计就变成为普通的最大似然估计．

（2）在J 取值于$[0,\infty)$时，常用广义分布$\frac{1}{J}$。

（3）在J 取值于$[0,1]$时，常用广义分布$\frac{1}{J(1-J)}$。

第二种取法是用最大熵原则：取所谓的最大熵分布（参见第 10 章），它是在已知的限制条件下，信息掌握得最少的分布，说一句白话， 就是“ 最吃亏的 ” 分布． 在对于参数完

全没有限制的情形下，广义的最大熵分布恰恰就是均匀分布． 所以第二种取法也是第一种取法的延伸。又若参数是一个概率向量（或者要估计的是一个取值于“概率向量组成的概率空间”的随机向量，这里

“n维概率向量组成的概率空间”$\overset{\Delta}{=}\{(p_1,\cdots,p_n):0\le p_1,\cdots,p_n\le 1, p_1+\cdots+p_n=1\}$），那么这时的最大熵分布是 Dirichlet 分布。这就是在生物信息论中常用 Dirichlet 分布作为先验分布的原因．

经验知识对于粗估先验分布常常很有用．例如，若未知的是英文字母或汉字，人们可以根据字母或汉字出现的经验频率得到先验分布的统计估计．

第三种取法是采用 Jeffreys 原则：取先验密度为 $f(q)=C\sqrt{\det I(q)}$， 其中C是常数，$I(q)$是信息矩阵，它的分量$I_{ij}(q)=-E(\frac{\partial^2 \ln p(x,q)}{\partial q_i \partial q_j}),(q=(q_1,\cdots,q_k))$．

第四种取法是统计中的所谓共轭分布方法．如果参数的先验分布与它关于分布$p(x,q)$的后验分布属于同一种类型，那么此种类型的分布就称为是$p(x,q)$的共轭分布．人们认为用共轭分布作为先验分布有其合理性，因为这时先验分布与后验分布属于同一种类型．

常见分布的共轭分布有： Beta 分布是二项分布的共轭分布，也是几何分布的共轭分布；Gamma 分布是 Poisson 分布的共轭分布，也是指数分布的共轭分布；正态分布是已知方差的正态分布的共轭分布；双侧截尾正态分布是已知方差的双侧截尾正态分布的共轭分布；逆 Gauss 分布是已知均值的正态分布（包括双侧截尾情形）的共轭分布；Perato 分布（$p(x)=g\cdot a^a x^{-(g+1)} I_{[a,\infty)}(x),(g,a>0)$）是均匀分布的共轭分布等．

在对参数没有任何先验知识时，在上面四种取法中，到底用那个更为合适？一般地说，并没有确切的答案．这是 Bayes 方法的不足之处．实际上这时只能随问题的不同作具体的考虑．

5．Bayes 估计再访

通过 Bayes 分布来估计参数或状态，有 3 种常见的方法：

（1）对给定的损失函数，求使后验风险最小的估计．

（2）用后验分布的数学期望，即后验期望估计．

（3）用后验分布达到最大值的点来估计，称为最大后验估计．

与一般数理统计类似地还可以用后验分布作区间估计．

6．Bayes 策略

定义９.３６ 在统计抉择中，通常会涉及样本，参数与行动．参数q是被考虑成随机变量的．将可能采取的行动全体组成的集合记为A．假定在对象的状态分布的参数为q时，如果采取行动$a(\in A)$，就会导致损失$L(q,a)$．而采取的动作a是依赖于被抽取到的样本$x_1,\cdots,x_n$的，我们将它记为 $a=a(\vec{x}),(\vec{x}=(x_1,\cdots,x_n))$．这个“函数”（从样本到动作的对应）$a(\bullet)$就称为一个策略。 在样本为$\vec{x}$时，策略$a(\vec{x})$带来的损失为$L(q,a(\vec{x}))$，它是样本的函数．

定义９.３７　在q固定的时的平均损失为

$$R(q,a(\bullet)) \overset{\Delta}{=} EL(q,a(\vec{x})) = \int L(q,a(\vec{x}))p(x_1,q)\cdots p(x_n,q)d\vec{x}\text{，} \tag{9. 44}$$

称为策略$a(\bullet)$关于q的风险函数。它对于先验分布$f(q)$的平均

$$r(a(\bullet)) \overset{\Delta}{=} \int R(q,a(\bullet))f(q)dq \tag{9. 45}$$

称为策略$a(\bullet)$对于先验分布f的 Bayes 风险。使先验分布f的 Bayes **风险达到最小的策略，**

称为对于先验分布f的 Bayes 策略。

定义９.３８　动作$a(\vec{x})$带来的损失$L(q,a(\vec{x}))$关于q的后验分布的期望，则称为策略

$a(\vec{x})$的后验风险（就是动作$a(\vec{x})$的后验风险），其表达式为

$$E_{Bayes}L(J,a(\vec{x})) = \int L(q,a(\vec{x}))p(q\,|\vec{x})dq\text{，} \tag{9. 46}$$

其中$p(q\,|\vec{x})$为参数q的后验密度，即后验分布的密度.

这里需要分清楚，策略的风险函数是对状态变量作的平均，即对于一切可能的样本作平均，所以它定义在整个策略类上，而策略对先验分布的 Bayes **风险，是策略的风险函数关于先验分布的平均．而后验风险是，在给定样本下，策略对参数的后验分布作的平均．从数学处理上，在求后者时样本是固定的，所以要简单得多．这就是使用后验分布的好处，当然，后验分布的计算也常常并不简单．**

需要注意，“ Bayes **”的冠名，通常用来修饰后验性质，例如，** Bayes **分布就是后验分布．但是，在** Bayes **风险的定义中，却被用来冠名先验性质，这似乎在冠名上有些混淆，但是因为有下面的定理，所以并无实质上的不妥.**

定理９．４０（Bayes 策略基本定理）

一个策略$a^*(\bullet)$是 Bayes **策略的充要条件是：对于任意样本$\vec{x}$，$a^*(\vec{x})$都使后验风险达到最小.**

（此定理的证明需要较深的知识，从略）.

这个定理说明了，Bayes **策略在任意样本$\vec{x}$上的取值，就是$\vec{x}$的最小后验风险动作．于是最小** Bayes **风险在本质上仍反映了后验性质.**

由此定理可见，只要算出了后验风险，再求得 Bayes **策略就并不困难。**

Bayes **估计可以纳入** Bayes **策略的框架．再者，假设检验也可以纳入** Bayes **策略的框架，因而** Bayes **统计方法的重点就归结为** Bayes **分布的计算。**

在实际问题中，Bayes **分布的计算，常常要通过随机模拟得到，特别是** Bayes **分布中的分母的计算，常常涉及巨大的工作量，较为有效的数值计算方法，就是** Markov **链** Monte Carlo **方法（**MCMC**）.**

4.２　Bayes **方法在图像中的应用与观测量不是状态变量时的参数估计**

图象处理包括，分割(Segmentation**)，图象压缩，图象重建，图象分类（如利用相似性对纹理图分类），图像识别，聚类等等.**

图象在传输过程先需要进行压缩，然后由接收方再将压缩图象信息恢复．所以，图象处理主要为：图象压缩与图象恢复．而图象分割常用在，例如，自然场景的分割．在图象恢复，图像分割的各种方法中，Bayes 方法是很重要的一个．它归结为在把接收到的信息作为已知的条件下，对描述图象的一些参数的重估.

Bayes 聚类方法，就是根据观测（从传输渠道得到)，在轮廓图(或灰度图)全体组成的集合中，选取一个在此观测的条件下，Bayes 风险最小的轮廓图(或灰度图)，作为恢复的图象．这里损失函数 $L(J,\hat{J})$ 的取法是十分重要的．它必须取得使估计的结果较为稳健 (robust)，即更多地与分布的定性行为相关，而对具体分布的表示式的敏感性尽量地小．这里不能一味简单地搬用模式识别中的现成结果，而必须考虑到"标的图象"的来源，结合它所在学科的特点，具体情形具体分析.

图像的先验分布 $p(a)$ (Priori distribution)，其中 a 为组态，一般可以写成 Gibbs 分布的形式(只要它处处为正):

$$p(a)=e^{-H(a)},$$

其中

$$H(a)=-\ln p(a),$$

称为能量函数或验前质量指标.

图像的信息特征中，一般还有代表边缘信息的 '边能量"（对应于 "边过程"），各部分的能量强度的不同常表现为某些加权参数 q_j 的大小．例如，图象恢复的模型可以取

$$H(a\mid b)=\sum_{j=1}^{3}q_jH_j(a\mid b\}\,.$$

这是一个参数建模问题，就是要确定参数，即要确定 H 为能量函数的 Gibbs 分布的参数 q_j 们． 在观测 $b^{(1)},\cdots,b^{(m)}$ （接收到的信息）条件下，作最大似然估计或 Bayes 估计．此时的联合分布(即似然函数)为

$$L(q_1,q_2,q_3)=\prod_{k=1}^{m}\frac{e^{-\sum_{j=1}^{3}q_jH(a|b^{(k)})}}{\sum_{g}e^{-\sum_{j=1}^{3}q_jH(g|b^{(k)})}}\,.$$

这里既有待估参数 $\vec{q}$，又有未知组态 a．要同时估计它们，一种有效的方法是将在第 15 章中介绍的 EM 算法，即用 EM 方法作迭代近似．对于格点数太多的情形，Geman 兄弟提出了一种简化的修正算法，可以大大地减少计算量．其基本思想为：一旦只要得到目标指向局部极大的 $\hat{q}^{(n)}$ 后，就按 $P(*\mid\hat{q}^{(n)},\vec{b})$ 取样 $\hat{a}^{(n)}$， 然后按下面的方法更新 $\hat{q}^{(n+1)}$ ：

任取 $\hat{q}^{(n+1)}$ 满足 $P(\hat{a}^{(n)}\mid\hat{q}^{(n+1)},\vec{b})>P(\hat{a}^{(n)}\mid\hat{q}^{(n)},\vec{b})$.

在实际估计中，对于参数的先验知识的利用是至关重要的，例如，参数所在的位置或区域.

Gibbs 场的 Bayes 分布纹理分析时，主要是根据纹理的特征，选取合理的 Gibbs 分布，即选取能量函数．常用的能量函数有 Graffigne 的$\Phi-$模型:

$$H(a)=\sum_{(x,y)\text{相邻}} q_{xy}\frac{1}{1+\dfrac{|a_x-a_y|^2}{d^2}}.$$

在模型选取好以后，纹理图的聚类就化为：通过观测到的一系列标准样品，进行学习训练以确定（估计）参数$\vec{q}$．再用 Bayes 聚类或其它聚类法来决定此纹理图形属于那一类.

图像识别的融合原则

在图象识别时，有时可以利用不同的识别器的融合 (fusion)，这时可以采取的原则，例如有，多数原则，权威原则，民主原则(以相似度作加权)等．

习题 9

1．设 w_n 独立同分布，且具有正连续密度 $r(x)$．又 $0<a\le g(x)\le b$，$x_{n+1}=f(x_n)+w_n(\mathrm{mod}\,1)$，$h_{n+1}=f(h_n)+w_n$．那么连续状态 Markov 链$x_n, h_n$都是指数 L_1 遍历的．

2．设 w_n 独立同分布，且具有正连续密度 $r(x)$．假定$\dfrac{\partial f}{\partial y}>0$，且存在 $g(y)\ge 0$，使在 x 固定时 $f(x,y)$ 对 y 的反函数，记为 $f^{-1}(x,y)$，满足 $r(f^{-1}(x,y)\ge \dfrac{\partial f(x,y)}{\partial y}g(y)$．而 Markov 链$x_n$满足 $x_{n+1}=f(x_n,w_n)$．那么x_n是指数L_1遍历的 Markov 链．

3．设 w_n 为独立同分布列，具有分布密度 r．证明

$$x_{n+1}=f(x_n,\cdots,x_{n-k+1})+w_n, x_0=x_0,\cdots,x_{k-1}=x_{k-1}$$

是k阶 Markov 链．意即

$$P(x_{n+k}\in A\,|\,x_n=x,x_{n-1}=x_{n-1},\cdots,x_0=x_0)=P(x_{n+k}\in A\,|\,x_n=x)\text{．}$$

龚光鲁，钱敏平著　应用随机过程教程 — 与在算法和智能计算中的应用
清华大学出版社，2003

第 10 章　隐马氏模型（Hidden Markov Model，HMM）及其应用

１． 熵与相对熵

１．1 离散分布的熵与相对熵

熵的概念出自 C.Shannon． 引进这个指标的目的在于刻画一个离散分布(一个离散随机变量)或一个分布密度(一个连续型随机变量)的不确定性的大小．也就是说知道了此随机变量的取值所获得的信息的大小．

定义１０．１　对于离散分布p$=(p_1,\cdots,p_n,\cdots)$，我们定义它的熵为

$$H(p)=-\sum_i p_i \ln p_i \text{ .}$$

又定义分布 p 关于分布 q$=(q_1,\cdots,q_n,\cdots)$的 Kullback- Leibler 相对熵为

$$h(p,q)=\sum_i p_i \ln \frac{p_i}{q_i}\text{ .}$$

命题１０．２　 相对熵 h（p,q）≥0，而且 h（p,q）=0，当且仅当 p = q 时成立等号．

证明　$[0,\infty)$上函数$g(t)\overset{D}{=}t-1-\ln t$　在$t\neq 1$时恒正（这一结论可由 g 的导数直接可以看出)，且$g(1)=0$． 取$t=\frac{q_i}{p_i}$，于是$\ln\frac{q_i}{p_i}$　≤$\frac{q_i}{p_i}$-1，即$\ln\frac{p_i}{q_i}$　$\geq 1-\frac{q_i}{p_i}$，而且等号仅当$\frac{p_i}{q_i}=1$时成立．从而有

$$\sum_i p_i \ln \frac{p_i}{q_i} \geq \sum_i p_i (1-\frac{q_i}{p_i})=0\text{ .}$$

所以结论成立.

这个命题表明，相对熵在相当程度上表达了 p 与 q 的差别：当 p = q 时，h（p,q）=0．而当所有的p_i都与q_i接近时，h（p,q）就很小．从而 h（p,q）可以看成 p 与 q 之间的一种“准距离”．这里我们之所以称它为准距离，是因为它既不对称（即 h（p,q）≠h（q,p）)，也不满足三角形不等式。所以不满足第 9 章中的距离公理.

例１０．３ （有限个值的分布的熵)

分布 p$=(p_1,\cdots,p_N)$的熵满足

$$H(p)=-\sum_i p_i \ln p_i \leq \ln N\text{ .}$$

且等号当且仅当分布在N个值均匀时成立．即以相同概率取N个值的分布(称为离散均匀分布)的熵最大．

证明　记分布 n =$(\frac{1}{N},\cdots,\frac{1}{N})$． 于是本结论是相对熵不等式 h（p,n）≥0 的变形.

例１０．４（数学期望固定条件下的离散的最大熵分布）

假定存在实数a，使$x_i \geq a,(i\geq 1)$．对于固定的$(x_1,\cdots,x_n,\cdots)$与m，在满足

$$P(x=x_i)=p_i, Ex=m$$

的离散分布 p= $(p_1,\cdots,p_n,\cdots)$ 中，具有最大熵的分布为 p^*，且

$$p_i^* = Ce^{-l_0 x_i}, (C = (\sum_i e^{-l_0 x_i})^{-1}),$$

其中 l_0 满足条件

$$\sum_i e^{-l_0 x_i} < \infty, \quad \sum_i x_i e^{-l_0 x_i} < \infty, \quad m\sum_i e^{-l_0 x_i} = \sum_i x_i e^{-l_0 x_i},$$

这个最大熵为 $ml_0 - \ln C$.

证明　对于任意 $l \neq 0$，只要 $\sum_i e^{-lx_i} < \infty$，相对熵不等式

$$0 \le h(p,p^*) = \sum_i p_i \ln \frac{p_i}{Ce^{-lx_i}}$$

$$= \sum_i p_i \ln p_i + l\sum_i p_i x_i - \ln C$$

就是

$$H(p) = -\sum_i p_i \ln p_i \le lm - \ln C .$$

所以

$$H(p^*) = -\sum_i Ce^{-l_0} \ln[Ce^{-l_0 x_i}] = l_0 m - \ln C \ge H(p).$$

例１０．５　(数学期望和方差都固定时的最大熵)

在均值为 μ，方差为 σ^2，且取值为 $x_k\ (k=1,2,\cdots)$ 的离散分布中，分布 p^*：

$$p_k^* = Ce^{-\frac{(x_k-a)^2}{b^2}}, C = (\sum_k e^{-\frac{(x_k-a)^2}{b^2}})^{-1}$$

的熵最大，此最大熵为 $\frac{s^2+(m-a)^2}{b^2} - \ln C$，其中 a, b 满足

$$\sum_k (x_k)^l e^{-\frac{(x_k-a)^2}{b^2}} < \infty, (l=0,1,2), \quad m\sum_k e^{-\frac{(x_k-a)^2}{b^2}} = \sum_k x_k e^{-\frac{(x_k-a)^2}{b^2}},$$

$$s^2(\sum_k e^{-\frac{(x_k-a)^2}{b^2}})^2 = [\sum_k e^{-\frac{(x_k-a)^2}{b^2}}][\sum_k x_k^2 e^{-\frac{(x_k-a)^2}{b^2}}] - (\sum_k x_k e^{-\frac{(x_k-a)^2}{b^2}})^2 .$$

(证明类似).

１．２ 分布密度的熵与相对熵

定义１０．６　对于概率分布密度，我们可以仿效前面的思想，把分布密度 $p(x)$ 的熵 (其实是微分熵)定义为

$$H(p) = -\int p(x)\ln p(x)dx .$$

而把分布密度 p(x)对分布密度 q(x)的相对熵定义为

$$h(p,q) = \int p(x)\ln\frac{p(x)}{q(x)}dx .$$

这个定义可以推广到多维密度 $p(x_1,\cdots,x_d)$ 与 $q(x_1,\cdots,x_d)$：

$$\text{H}(p)=\int p(x_1,\cdots,x_d)\ln\frac{p(x_1,\cdots,x_d)}{q(x_1,\cdots,x_d)}dx_1\cdots dx_d\ ),$$

$$h(p,q)=\int p(x_1,\cdots,x_d)\ln\frac{p(x_1,\cdots,x_d)}{q(x_1,\cdots,x_d)}dx_1\cdots dx_d\ .$$

同样也有相对熵是非负的．　从而由熵的定义，再利用相对熵非负性质可推得下述结论：

例１０．７（半直线上均值已知时的最大熵分布密度）

在 $[0,\infty)$ 上具有同均值 m 的密度中，指数分布 $Exp_{\frac{1}{m}}$ 的熵最大．这个最大熵为 $1+\log m$ ．

例１０．８　(均值与方差已知的分布密度的最大熵)

在均值为 m ，方差为 s^2 的分布密度中，正态分布的熵最大．

例１０．９（均值向量与(协)方差矩阵已知的有联合密度的最大熵)

在均值向量为 $\vec{m}$ ，方差矩阵为 Σ 的多维密度中，Gauss 分布 $N(\vec{m},\Sigma)$ 的熵最大．请读者自己写出这个最大熵的表达式．

例１０．１０　（有限区间上的最大上密度）

在有限区间 $[a,b]$ 上取值的分布密度中，均匀分布的熵最大．此最大熵为 $\ln(b-a)$ ．

例１０．１０也可以推广到多维情形．

例１０．１１　（在取值于混合概率向量的分布密度中，关于某个概率向量的平均相对熵给定条件下，求具最大相对熵的密度）

令样本空间为

$$\Omega=\{\ \vec{x}=(x_1,\cdots,x_d):0<x_i<1,(i\le d),x_1+\cdots+x_d=1\}\ .$$

Ω 中的点可以解释为概率向量，而取值于 Ω 的连续型随机变量的分布密度可以解释为 ”广义的混合分布 ”．给定 $\vec{z}=(z_1,\cdots,z_d)\in W$ ，把 Ω 中的点看成概率向量，就有相对熵 $h(\vec{z},\vec{x})$ ，它是取值于 Ω 的 $\vec{x}$ 函数．对于一个取值 Ω 的分布密度 $p(\vec{x})\overset{\Delta}{=}p(x_x,\cdots,x_d)$ 定义对于这个分布密度的平均相对熵：

$$m(\vec{z})\overset{D}{=}\int h(\vec{z},\vec{x})p(\vec{x})d\vec{x}\ .$$

我们要在平均相对熵 $m(\vec{z})$ 相等的分布密度 $p(\vec{x})$ 中，选取一个 $p^*(\vec{x})$ ，使其熵 $-\int p^*(\vec{x})\ln p^*(\vec{x})d\vec{x}$ 达到最大．

为此我们断言：如下的 Gibbs 型分布密度

$$p^*(\vec{x})=C(\vec{z})\,e^{-l(\dot{z})h(\dot{z},\dot{x})}\ (=C(\vec{z})\,e^{-l(\dot{z})\sum_i z_i\ln z_i}x_1^{l(\dot{z})z_1}\cdots x_d^{l(\dot{z})z_d})\ ,$$

其中

$$C(\vec{z})=(\int e^{-l(\dot{z})h(\dot{z},\dot{x})}d\vec{x})^{-1}\ ,$$

$l(\vec{z})$ 满足

$$m(\vec{z})=C(\vec{z})\int h(\vec{z},\vec{x})e^{-l(\dot{z})h(\dot{z},\dot{x})}d\vec{x}\ ,$$

就是我们要的具有最大熵的密度．证明如下：对于任意一个满足 $m(\vec{z})\overset{D}{=}\int h(\vec{z},\vec{x})p(\vec{x})d\vec{x}$ 的分布密度 $p(\vec{x})$ ．由相对熵不等式

$$0 \le h(p, p^*) = \int p(\vec{x}) \ln \frac{p(\vec{x})}{C(\vec{z}) e^{-l(\vec{z}) h(\vec{z},\vec{x})}} d\vec{x}$$

$$= -H(p) + l(\vec{z}) m(\vec{z}) - \ln C(\vec{z})$$

得到

$$H(p) \le l(\vec{z}) m(\vec{z}) - \ln C(\vec{z}) .$$

于是我们有

$$H(p^*) = -\int C(\vec{z}) e^{-l(\vec{z}) h(\vec{z},\vec{x})} \ln[C(\vec{z}) e^{-l(\vec{z}) h(\vec{z},\vec{x})}] d\vec{x} = l(\vec{z}) m(\vec{z}) - \ln C(\vec{z}) \ge H(p) .$$

［注 1］上面我们用相对熵 $h(\vec{z},\vec{x})$ 作为准距离，而不是用更为合理的 $\frac{1}{2}[h(\vec{z},\vec{x}) + h(\vec{x},\vec{z})]$. 也可以用与它相应的 $h(\vec{x},\vec{z})$ ，只要易于计算，不管用那一个都行.

［注２］ Diriclet 分布的密度函数为

$$f(x_1, \cdots, x_d) = C x_1^{a_1 - 1} \cdots x_d^{a_d - 1} I_{\{x_1 + \cdots + x_d = 1\}} ,$$

其中

$$C = \frac{\Gamma(a_1) \cdots \Gamma(a_d)}{\Gamma(a_1 + \cdots + a_d)}, \Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx .$$

例１０.１１中的最大熵分布恰是 Dirichlet 分布. 在生物信息论中，常常需要估计概率向量组成的空间上的分布密度. 例１０.１１说明，在平均相对熵给定的条件下，Dirichlet 分布是"最吃亏的"分布. 用它作为先验分布"看起来更为保险". 这就解释了在生物信息论中，人们常常喜欢用 Dirichlet 分布作为先验分布的原因.

相对熵应用的一个实例 — 在各个测试特征（统计量）中选择几个最为有效的相对熵方法（特征量选取的相对熵方法)

假定我们有两组性质完全不同的群体， 例如一组是健康人，另一组是 SARS （非典型性肺炎）病毒持带人. 又假定人们已经提出了 N 种区别健康人与 SARS 病毒持带人的不同特征. 我们要在其中选取区别效果最好的 M 个特征. 相对熵方法是较为有效的一种方法，其实际操作为:

用一个给定的区分特征，对上面的一组健康人测定了一组数据，简称为甲数据组；又对上面的 SARS (Severe Acute Respiratory Syndrome) 病毒持带人测定了一组数据，简称为乙数据组. 再进行如下的步骤：

（1）分别找出此两个数据组的近似分布：从数据组出发，应用统计中的核估计的思想，分别得到其近似分布密度曲线；

（2）计算此两个分布密度的相对熵 h_0 ；

（实践证明，用数值计算求此两个分布密度的相对熵，对于计算格点大小的划分并不太敏感. 相反地，如果直接将两个直方图作为离散分布求相对熵，则对于直方图的计算格点大小的划分十分敏感，以致得到的计算结果很不稳定）.

（3）将甲乙两组数据合并,再随机地重组为和以前个数相同的两组(随机地重排（Random Sorting）后，再按原来的各组的个数顺序分成两组)，用（１），（２）步骤计算其相对熵；

（4）重复地作（3）多次（例如１万次），计算其中相对熵大于 h_0 的次数所占的频率，记为 p .

于是对应于一个区分特征就对应地有一个 p 值，它是一个客观的标准，和各种测试方法的结果（统计量）的绝对大小、量纲都无关．如果要在 N 个不同的用以区分是否感染了SARS 的特征中选取 M 个相对地最为有效的特征，我们只需选取对应的 p 值最小的前 M 个特征即可．

2．隐 Markov 模型（HMM）

2．1　一个实例

先举一个例子以直观地理解这个模型的实质.

例１０.１２　设某人在三个装有红白两种颜色的球的盒子中，任取一个盒子，然后在此盒子中每次抽取一个球，连续地抽取 m 次．假定各个盒子的内容与抽取方式分别为：

	红球数	白球数	每次抽取方式
盒１	90	10	随机取一球，记下颜色后不放回，而放进一个与它颜色不同的球．
盒２	50	50	随机取一球，记下颜色后放回
盒３	40	60	随机取一球，记下颜色后不放回，并放进一个红球．

现在如果某人用上述方法得到了一个记录(红, 红, 红, 红, 白)（即$m=5$），但是不告诉我们球出自哪个盒子，我们应如何推测他是从哪个盒子中抽取的观测样本呢 ？

令

$$S_n^{(k)} = \text{在第}k\text{个盒子}\ (k=1,2,3)\ \text{中第 n 次抽取完成后在各盒中的红球数}.$$

那么，在 k 分别固定为１，２，３时，$\{S_n^{(k)}: n\ge 0\}$ 分别为 Markov 链，且其概率转移矩阵分别为：

$$p_{ij}^{(1)}=\begin{cases}\dfrac{i}{100} & (j=i-1)\\ 1-\dfrac{i}{100} & (j=i+1)\\ 0 & (\text{其它情形})\end{cases}，\text{（逢红，红减１；逢白，红加１）},$$

$$p_{ij}^{(2)}=\begin{cases}1 & (j=i)\\ 0 & (j\ne i)\end{cases}，\text{（内容总是不变）}$$

$$p_{ij}^{(3)}=\begin{cases}\dfrac{i}{100} & (j=i)\\ 1-\dfrac{i}{100} & (j=i+1)\\ 0 & (\text{其它情形})\end{cases}．\text{（逢红不变；逢白，红加１）}$$

而且初值分别为：$S_0^{(1)}=90, S_0^{(2)}=50, S_0^{(3)}=40$．于是这３个盒子就分别对应于３个不同的 Markov 链模型．把这３个模型分别简记为 l_1, l_2, l_3，并把某人观测到的样本序列中的第 n 个记为 O_n．即令

$$O_n = \text{抽到的记录列中第}n\text{个记录中的白球数（只能为 1 或 0）}.$$

此例是一个玩具模型．从这个例子可以看出，在观测出自哪个盒子已知时，状态随机变

量序列$\{S_n\}$与某人提供的观测随机变量序列$\{O_n\}$之间的条件概率计算的关系可以直观地写为：

$$S_0, O_1, S_1, \cdots O_{m-i}, S_{m-1}, O_m, S_m,$$

其中在前面的一段随机变量序列取定值的条件下，继后的那个随机变量取值的条件概率就完全确定了．而且在这３个模型下分别都有

$$P(S_{n+1} \mid O_1, S_1, \cdots O_n, S_n) = P(S_{n+1} \mid S_n) = P(S_2 \mid S_1),$$

$$P(O_{n+1} \mid O_1, S_1, \cdots O_{m-i}, S_{m-1}, O_m, S_m) = P(O_{n+1} \mid S_n) = P(O_2 \mid S_1).$$

于是

$$P(S_0, O_1, S_1, \cdots O_{m-i}, S_{m-1}, O_m, S_m)$$
$$= P(S_0)P(O_1 \mid S_0)P(S_1 \mid S_0)P(O_2 \mid S_1)\cdots P(O_m \mid S_{m-1})P(S_m \mid S_{m-1}).$$

具体地，我们有

在模型 l_1 下（把取到的球换色）

$$P((O_1, O_2, O_3, O_4, O_5) = (0,0,0,0,1) \mid l_1) = 0.9 \times 0.89 \times 0.88 \times 0.87 \times 0.13 \approx 0.08.$$

在模型 l_2 下（球的内容不变），O_n 是独立同分布的随机变量序列，且 $O_n \sim \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$，

所以

$$P((O_1, O_2, O_3, O_4, O_5) = (0,0,0,0,1) \mid l_2) = (\frac{1}{2})^5 \approx 0.03.$$

在模型 l_3 下（取到红则不变，取到白则换为红球）

$$P((O_1, O_2, O_3, O_4, O_5) = (0,0,0,0,1) \mid l_3) = (0.4)^4 \times 0.6 \approx 0.015.$$

再用Bayes 公式分别得到（即取上面３ 个概率的归一化值）

$$P(l_1 \mid (O_1, O_2, O_3, O_4, O_5) = (0,0,0,0,1)) = 0.64,$$

$$P(l_2 \mid (O_1, O_2, O_3, O_4, O_5) = (0,0,0,0,1)) = 0.24$$

$$P(l_3 \mid (O_1, O_2, O_3, O_4, O_5) = (0,0,0,0,1)) = 0.12.$$

可见从第一盒抽出样本(红，红，红，红，白)的概率要比从其他两盒中抽出该样本的概率要大得多．

这个模型的意义绝对不在于游戏，从中可以抽象出一种能广泛应用的数学模型－隐Markov 模型 （Hidden Markov Model， 简记为 HMM）．在这个例子中的不同盒的抽取方式可以抽象为不同的编码方式． 而这个例子正启示了用隐 Markov 模型作模式识别的一种粗略梗概.

２．２ 隐 Markov 模型的描述

定义１０．１３ 设$\{X_n : n \geq 1\}$是取值于有限状态$\{1, \cdots, L\}$的随机变量列，称为状态链，X_1的分布称为其初始分布，记为$\vec{m}$．假定 X_n 是时齐的 Markov 链．记

$$a_{ij} = P(X_{n+1} = j \mid X_n = i), \quad A = (a_{ij})_{i,j \leq L}. \qquad (10.1)$$

为它的转移矩阵．假定 X_n 的取值，初始分布$\vec{m}$与转移矩阵A，都不能测量得到．而能测量到的是另一个与它有联系的，且可以观测到的一个取值于与有限集 $\{n_1, \cdots, n_M\}$的随机变量序列Y_n，称为观测链，它们合起来还要求满足如下的隐 Markov 条件（HMM 条件）：

$$P(\vec{Y}=\vec{y},\vec{X}=\vec{x})=m_{i_1}b_{i_1y_1}a_{i_1i_2}\cdots b_{i_{N-1}y_{N-1}}a_{i_{N-1}i_N}b_{i_Ny_N},\qquad(10.2)$$

其中 N 为样本观测的时间长度，而 $\vec{X}=(X_1,\cdots,X_N)$，$\vec{Y}=(Y_1,\cdots,Y_N)$，$\vec{y}=(y_1,\cdots,y_N)$，$\vec{x}=(i_1,\cdots,i_N)$，$y_n\in\{n_1,\cdots n_M\},1\le i_n\le L,1\le n\le N$，初始概率向量（初始分布）为 $\vec{m}=(m_1,\cdots,m_L)$．由未知状态链与测量到的观测链一起(X_n,Y_n)，就构成了隐 Markov 模型 ，这里“隐”的含义是说状态链是隐藏起来的．

隐 Markov 模型的基本假定是：

参数向量$\vec{m}$，参数矩阵 A 与 $B\overset{\Delta}{=}(b_{il})_{L\times M}$ （ $i\in\{1,\cdots,L\},l\in\{n_1,\cdots,n_M\}$）都是未知的，我们将它们合记为参数组 $\vec{l}\overset{\Delta}{=}(\vec{m},A,B)$.参数组 $\vec{l}\overset{\Delta}{=}(\vec{m},A,B)$ 完全确定了状态链与观测连的联合统计规律，所以,我们通常用一个$\vec{l}$ 表示一个隐 Markov 模型，并称之为隐 Markov 模型 $\vec{l}$ (更确切地为隐 Markov 链)． 在例 10.12 中，３个不同的模型就对应了３个不同的参数组．只要令 $X_n=S_n,Y_n=O_{n+1}$，它们满足 HMM 条件，因而纳入了隐 Markov 模型的框架．

(10.2)是$(\vec{X},\vec{Y})$ 的联合分布通过参数表达的形式，它是计算各种边缘概率与条件概率的出发点．而 HMM 条件的含义是：状态链与观测链的联合分布是由一系列简单转移与条件概率的乘积表达的．

２．３　隐 Markov 模型的等价表述

由例１０．１２的启示，可以想到下面的等价条件

命题１０．１４

HMM 条件等价于：对于任意的$i\in(1,\cdots,L\}$ 以及 $l\in\{n_1,\cdots,n_M\}$，有

$$P(Y_n=v_l\mid X_n=i,Y_{n-1}=i_{n-1},X_{n-1}=n_{l_{n-1}},\cdots,Y_1=i_1,X_1=n_{l_1})=P(Y_n=v_l\mid X_n=i)\overset{D}{=}b_{il},\qquad(10.3)$$

$$P(X_{n+1}=j\mid X_n=i,Y_n=n_{l_n},X_{n-1}=i_{n-1},\cdots,X_1=i_1,Y_1=n_{l_1})=P(X_{n+1}=j\mid X_n=i)\overset{\Delta}{=}a_{ij},\qquad(10.4)$$

这两个等式的证明只需利用条件概率的定义．所以我们把它留给读者．它们的直观含义就是：Y_n 与 X_{n+1} 相对于历史条件的统计规律只与时间上最接近的 X_n 有关， 而与其它更早的历史无关．在实际问题中， 这是比较容易判断的．

２．４　非线性滤波作为隐 Markov 模型的特例

设 X_n 是取值于状态空间$S=\{1,2,\cdots,L\}$ 的 Markov 链，其样本不能被实际测量得到．而能测量到的是如下的Y_n 的样本

$$Y_n=g(X_n)+w_n,\qquad(10.5)$$

其中g 是一个未知函数，$\{w_n\}$是独立同分布的随机干扰， 只取有限个值， 且与随机过程$\{X_n\}$独立．那么 $\{X_n,Y_n\}$就是一个（二维的）隐 Markov 模型．这就把（１０．５）的非线性滤波放进了 HMM 的框架．

２．５　在应用中研究隐 Markov 模型的主要方面

(1) 从一段观测序列$\{Y_k,k\le m\}$及已知模型$\vec{l}=(\vec{m},A,B)$ 出发，估计 X_n 的最佳值，称为解码问题．这是状态估计问题．

(2) 从一段观测序列$\{Y_k, k \le m\}$出发，估计模型参数组 $\vec{l} = (\vec{m}, A, B)$，称为学习问题．就是参数估计问题．

(3) 对于一个特定的观测链$\{Y_k, k \le m\}$，已知它可能是由已经学习好的若干模型之一所得的观测，要决定此观测究竟是得自其中哪一个模型．这称为识别问题．就是分类问题．

在实际问题中，这 3 个问题有时并不完全能分开．有时也并不需要解全 3 个问题．例如，在语音识别或手写体汉字或数字的脱机识别中，我们只需要作(2)与(3)，这相当于将一个 ”标准的” 语音音素或一个 ”标准的” 手写体汉字 “学习成” 一个隐 Markov 模型，即把它与一个或几个特定的隐 Markov 模型(更确切地说，是特定的一组参数)相对应起来，以便把该模型作为这个音素或手写字的代表模板，这是学习相位；而在进一步用这些模板中的最合适者，作为对于一个需要识别的音素或手写字的分类归属，这是运转相位．至于归入那个模板最合适，就要用合理的距离，或准距离（常用的是相对熵），在此意义下优化．

[注 1] X_n 也可以推广为：取值于平面有限格点的 Markov 场，而Y_n 与它的关系仍如上述，这就定义了一个隐 Markov 场 ．

[注 2] 一般地，Y_n 还可以是连续型随机变量． 如果记在 $X_n = x$ 的条件下，Y_n 的条件分布密度为 f，则$\vec{l} = (\vec{m}, A, f_x)$ 就也称为一个（连续的）隐 Markov 模型． 这时 f 常是分布类型已知而带有未知参数的密度．状态是连续的隐 Markov 模型至今还未见有人使用．

3 解码问题 -- 已知模型 $\vec{l}$ 与观测 $\vec{Y} = \vec{y}$ 时状态 $\vec{X}$ 的估计

3．1 出现当前的观测的概率 $P(\vec{Y} = \vec{y} \mid \vec{l})$ 的计算

我们仍旧沿用记号

$$\vec{X} = (X_1, \cdots, X_N),\ \vec{Y} = (Y_1, \cdots, Y_N),\ \vec{y} = (y_1, \cdots, y_N),\ \vec{x} = (i_1, \cdots, i_N),$$

$y_n \in \{n_1, \cdots n_M\}, 1 \le i_n \le L, 1 \le n \le N$，初始概率向量 $\vec{m} = (m_1, \cdots, m_N)$．

由(10．2)，利用条件概率的性质容易算出

$$P(\vec{Y} = \vec{y} \mid \vec{X} = (i_1, \cdots, i_N), \vec{l}) = b_{i_1 y_1} \cdots b_{i_N y_N},$$

$$P(\vec{Y} = \vec{y}, \vec{X} = \vec{x} \mid \vec{l}) = m_{i_1} b_{i_1 y_1} a_{i_1 i_2} \cdots b_{y_{N-1} y_N} a_{i_{N-1} i_N} b_{i_N y_N},$$

$$P(\vec{Y} = \vec{y} \mid \vec{l}) = \sum_{i_1, \cdots, i_N} m_{i_1} b_{i_1 y_1} a_{i_1 i_2} \cdots a_{i_{N-1} i_N} b_{i_N y_N}. \qquad (10．6)$$

对于 $1 \le n \le N$ 及观测样值 $\vec{Y} = \vec{y}$，记（因为观测样值 $\vec{y}$ 是固定的，所以下面我们将在足标中把它略去）

$$a_n(i) = P(Y_1 = y_1, \cdots, Y_n = y_n, X_n = i \mid \vec{l})\ (\text{依赖 } \vec{y}), \qquad (10．7)$$

则在模型$\vec{l}$ 给定下，关于观测资料$(y_1, \cdots, y_n)$的长度 $n\ (n < N)$，我们有递推公式(称为向前递推公式或向前算法)

$$a_{n+1}(i) = \sum_j a_n(j) a_{ji} b_{i y_{n+1}}. \qquad (10．8)$$

由此得到

(1) 基于$a_n(i)$ 的向前递推公式计算 $P(\vec{Y} = \vec{y} \mid \vec{l})$ 的步骤：

计算初值 $a_1(i) = m_i b_{i y_1}$，

用递推公式 $a_{n+1}(i)=\sum_j a_n(j)a_{ji}b_{iy_{n+1}}$， （10．9）

最后得到结论 $P(\vec{Y}=\vec{y}\mid\vec{l})=\sum_i a_N(i)$． （10．10）

计算 $P(\vec{Y}=\vec{y}\mid\vec{l})$ 也可以通过另一种途径，为此记（观测样值 $\vec{y}$ 也是是固定的,我们也把它在足标中略去）

$$b_n(i)=P((Y_{n+1}=y_{n+1},\cdots,Y_N=y_N\mid X_n=i,\vec{l})\,. \qquad (10．11)$$

则在模型 $\vec{l}$ 给定下，关于观测资料 $(y_1,\cdots,y_n)$ 的长度 $n\ (n<N)$，我们还有递推公式(称为向后递推公式或向后算法)

$$b_n(i)=\sum_j b_{n+1}(j)a_{ij}b_{jy_{n+1}}$$

这就得到计算 $P(\vec{Y}=\vec{y}\mid\vec{l})$ 的另一个算法，即

(2) 基于 $b_n(i)$ 的向后递推公式计算 $P(\vec{Y}=\vec{y}\mid\vec{l})$ 的步骤

定义 $b_N(i)=1$，

用递推公式 $b_n(i)=\sum_j b_{n+1}(j)a_{ij}b_{jy_{n+1}}$ （10．12）

最后得到结论 $P(\vec{Y}=\vec{y}\mid\vec{l})=\sum_i b_0(i)m_i b_{iy_1}$ （10．13）

3．2 解码问题 -- 已知模型 $\vec{l}$ 与观测 $\vec{Y}=\vec{y}$ 时状态 $\vec{X}$ 的估计

令

$$g_n(i)=P(X_n=i\mid\vec{Y}=\vec{y},\vec{l})\,, \qquad (10．14)$$

那么

$$g_n(i)=\frac{P(\vec{Y}=\vec{y},X_n=i\mid\vec{l})}{\sum_i P(\vec{Y}=\vec{y},X_n=i\mid\vec{l})}=\frac{a_n(i)b_n(i)}{\sum_i a_n(i)b_n(i)}\,. \qquad (10．15)$$

(推导的思路如下： 令 $\mathrm{F}_n=s(X_m,Y_m,m\le n)$ (右方表示括号内随机变量所生成的 s 代数)，$Y_n^-=\{Y_1=y_1,\cdots,Y_n=y_n\}$，$Y_n^+=\{Y_{n+1}=y_{n+1},\cdots,Y_N=y_N\}$．因为 $\vec{l}$ 是固定的，我们在下面记号中也把它（为了记号的方便，我们在下面有时并不区分取概率的符号与取期望的符号,取概率也理解为取期望）．于是我们有

$$P(\vec{Y}=\vec{y},X_n=i\mid\vec{l})=E[(P(\vec{Y}=\vec{y},X_n=i\mid\mathrm{F}_n)]=E[Y_n^-,X_n=i,P(Y_n^+\mid X_n=i,\ \mathrm{F}_n)]$$
$$=E[Y_n^-,X_n=i,P(Y_n^+\mid X_n=i)]=P(Y_n^-,X_n=i)P(Y_n^+\mid X_n=i)=a_n(i)b_n(i)\,).$$

对于解码问题的解决，我们用如下两种考虑方法:

1． 对单个时刻的状态的最大概率估计

如果 i^* 满足 $g_n(i^*)=\max_{i\le N} g_n(i)$，则取 $\hat{X}_n=i^*$．

单个时刻的状态的最大概率估计的长处在于简单易算．但是它忽视了状态链在不同时刻之间的联系，也就不能充分地利用已知的信息．下面的路径估计法，是在考虑不同时间的联系后的改进方法．

2　路径的(轨道)最大概率估计 -- Viterbi 算法

这是基于动态规划的思想的一种整体性考虑的方法（所谓动态规划实质上就是一种从后向前逐步取优的优化方法）．令在时刻 n 取到状态 i 的各个路径的最大概率为

$$d_n(i) = \max_{i_1,\cdots,i_{n-1}} P(X_n = i, X_{n-1} = i_{n-1}, \cdots, X_1 = i_1, Y_n = y_n, \cdots, Y_1 = y_1 \mid \vec{l}), \quad (10.16)$$

那么我们有

$$d_{n+1}(i) = b_{iy_{n+1}} \max_j (d_n(j) a_{ji}). \quad (10.17)$$

用来计算它的实际程式是如下的 Viterbi 算法：假定最佳路径为 $(\hat{X}_1, \cdots, \hat{X}_N)$．今设置一个从 $\hat{X}_{n+1}$ 得到 $\hat{X}_n$ 的对应 $y_{n+1}: \hat{X}_n = y_{n+1}(\hat{X}_{n+1})$．计算 $d_n(i)$ 与 y_n 的递推算法的具体设计如下：

设置初值　$d_1(i) = m_i b_{iy_1}$；

用递推公式　$d_{n+1}(i) = b_{iy_{n+1}} \max_j (d_n(j) a_{ji})$；

再定义 $y_{n+1}(i) = j^*$，其中 j^* 只要满足

$$d_{n+1}(i) = b_{iy_{n+1}} d_n(j^*) a_{j^*i}. \quad (10.18)$$

最终采取如下的向后递推算法（相当于动态规划， 动态规划的核心思想是：在一切最佳路径上的任意节点出发，其后也应该是最佳路径， 因此就能从后向前地逐步求得最佳路径）得到如下的“最佳”轨道估计 $(\hat{X}_1, \cdots, \hat{X}_N)$：

计算终值　$\hat{X}_N = i^* \in \{i : d_N(i) = \max_j d_N(j)\}$，　(10.19)

反向归纳地取　$\hat{X}_n = y_{n+1}(\hat{X}_{n+1})$．　(10.20)

4　学习问题 – 由观测 $\vec{Y} = \vec{y}$ 估计模型参数 $\vec{l}$

4.1　状态链样本已知时的参数的频率估计

设隐 Markov 模型的状态链有充分长的样本．把其中从状态 i 到下一个时刻为状态 j 转移的频数记为 A_{ij}．那么，粗略地可取 $\hat{a}_{ij} = \dfrac{A_{ij}}{\sum_{j=1}^{N} A_{ij}}$ 为 a_{ij} 的估计. 同样设此观测链样本充分长, 记其中从状态 i 到同一时刻的观测 l 转移的频数为 B_{ij}，则可取 $\hat{b}_{il} = \dfrac{B_{il}}{\sum_{l=1}^{M} B_{il}}$ 为 b_{il} 的估计．不幸的是状态链是经过估计得到的， 所以不修正地使用频率估计会增加误差. 再则, 这种估计也过于敏感.

4．2　模型参数估计的 EM 算法的思想

我们取 $\vec{l}$ 的最大似然估计 $\hat{l}$，即它满足

$$P(\vec{Y} = \vec{y} \mid \hat{l}) = \max_{\vec{l}} P(\vec{Y} = \vec{y} \mid \vec{l}) \quad (10.21)$$

但是，对于似然函数$P(\vec{Y}=\vec{y}\mid\vec{l})=\sum_{i_1,\cdots,i_N} m_{i_1} b_{i_1 y_1} a_{i_1 i_2}\cdots a_{i_{N-1} i_N} b_{i_N y_N}$而言， 这个最大值问题的计算量过大，在实际中是不可能被采用的．为此，人们灵活地利用使似然函数尽量地大的思想，采取折衷的方案，构造一个递推算法，使之能相当合理地给出模型参数$\vec{l}$的粗略估计．其核心思想是： 并不要求备选估计$\vec{l}$使$P(\vec{Y}=\vec{y}\mid\vec{l})$（关于$\vec{l}$）达到最大或局部极大，而只要求使$P(\vec{Y}=\vec{y}\mid\vec{l})$相当大．从而计算变为实际可能．

为此，我们定义一个描述模型“趋势”的量 $Q(\vec{l}^*\mid\vec{l})$代替似然函数$P(\vec{Y}=\vec{y}\mid\vec{l})$，其定义为

$$Q(\vec{l}^*\mid\vec{l})\overset{\Delta}{=}\sum_{\vec{x}} P(\vec{X}=\vec{x},\vec{Y}=\vec{y}\mid\vec{l})\ln P(\vec{X}=\vec{x},\vec{Y}=\vec{y}\mid\vec{l}^*)$$

$$\overset{\text{简记为}}{=}\sum_{\vec{x}} P(\vec{x},\vec{y}\mid\vec{l})\ln P(\vec{x},\vec{y}\mid\vec{l}^*) \qquad (10.22)$$

(这个量并不是相对熵,但是在形式上很类似)．利用在$0<x<1$时的不等式$\ln x\le x-1$

(证明：令 $f(x)=(x-1)-\ln x$，则 $f(1)=0$ 且 $f''(x)=\frac{1}{x^2}>0$．即$f(x)$为凸函数，从而 $f(x)\ge 0$）得到

$$Q(\vec{l}^*\mid\vec{l})-Q(\vec{l}\mid\vec{l})=\sum_{\vec{x}} P(\vec{x},\vec{y}\mid\vec{l})\ln\frac{P(\vec{x},\vec{y}\mid\vec{l}^*)}{P(\vec{x},\vec{y}\mid\vec{l})}$$

$$\le\sum_{\vec{x}} P(\vec{x},\vec{y}\mid\vec{l})\left(\frac{P(\vec{x},\vec{y}\mid\vec{l}^*)}{P(\vec{x},\vec{y}\mid\vec{l})}-1\right)=P(\vec{Y}=\vec{y}\mid\vec{l}^*)-P(\vec{Y}=\vec{y}\mid\vec{l})\,.$$

由此可见对于固定的$\vec{l}$，只要 $Q(\vec{l}^*\mid\vec{l})>Q(\vec{l}\mid\vec{l})$，就有$P(\vec{Y}=\vec{y}\mid\vec{l}^*)>P(\vec{Y}=\vec{y}\mid\vec{l})$．于是想把模型$\vec{l}_n$修改为更好的模型$\vec{l}_{n+1}$（即找$\vec{l}_{n+1}$使 $P(\vec{Y}=\vec{y}\mid\vec{l}_{n+1})>P(\vec{Y}=\vec{y}\mid\vec{l}_n)$），只需找$\vec{l}_{n+1}$使

$$Q(\vec{l}_{n+1}\mid\vec{l}_n)=\sup_{\vec{l}} Q(\vec{l}\mid\vec{l}_n), \qquad (10.23)$$

即只要把$Q(\vec{l}\mid\vec{l}_n)$关于$\vec{l}$的最大值处取成$\vec{l}_{n+1}$，就有$P(\vec{Y}=\vec{y}\mid\vec{l}_{n+1})>P(\vec{Y}=\vec{y}\mid\vec{l}_n)$．

这样得到的模型序列$\{\vec{l}_n\}$能保了$P(\vec{Y}=\vec{y}\mid\vec{l}_n)$关于$n$是严格递增的．虽然在这里还不能在理论上证明$P(\vec{Y}=\vec{y}\mid\vec{l}_n)$收敛到$\max_{l} P(\vec{Y}=\vec{y}\mid\vec{l})$，但是当$n$充分大时，$\vec{l}_n$也还能提供在实际中较为满意的粗略近似．

综上论述， 我们把如上得到近似模型列$\vec{l}_n$的方法归结为两个步骤：

1 E 步骤（求期望）：计算

$$Q(\vec{l}^*\mid\vec{l})=\sum_{\vec{x}} P(\vec{x},\vec{y}\mid\vec{l})\ln P(\vec{x},\vec{y}\mid\vec{l}^*)\ ;$$

2 M 步骤 （求最大）：求$\vec{l}_{n+1}$使

$$Q(\vec{l}_{n+1}\mid\vec{l}_n)=\sup_{\vec{l}} Q(\vec{l}\mid\vec{l}_n)。$$

这两个步骤合起来构成的算法，称为 EM 算法．EM 算法是针对在测量数据不完全时，求参数的一种近似于最大似然估计的统计方法．隐 Markov 模型的建模（模型参数的估计），是 EM 算法的一个最常见，且极有用的一种典型例子．下面将看到，隐 Markov 模型中的 M – 步骤的解可以有显式表示，这就是一组把模型参数修改为新的模型参数的递推公式．这组公式正好是在隐 Markov 模型中普遍应用的著名的 Welch-Baum 公式．下面我们给出得到这组递推公式的梗概．首先我们有下面的引理．

引理１０．１５　设 $z_i \ge 0 \quad (i \le N)$，则在约束条件 $\sum_{i=1}^{N} x_i = 1$，$x_i \ge 0 \quad (i \le N)$ 下，函数 $\sum_{i=1}^{N} z_i \ln x_i$ 在 $x_i = \frac{z_i}{\sum_{i=1}^{N} z_i} \quad (i \le N)$ 处取得最大．

证明　注意 $\frac{1}{\sum_{j=1}^{N} z_j}(z_1, \cdots, z_N)$ 对 $(x_1, \cdots, x_N)$ 的相对熵 $\sum_{i=1}^{K}\left(\frac{z_i}{\sum_{j=1}^{K} z_j}\right)\ln\frac{\left(\frac{z_i}{\sum_{j=1}^{K} z_j}\right)}{x_i} \ge 0$．

４．３　隐 Markov 模型中 M – 步骤的求解

现在我们用引理 10．1，来求 $Q(\vec{l} \mid \vec{l}_n)$ 关于 $\vec{l}$ 的最大值的位置 $\vec{l}_{n+1}$．注意

$$P(\vec{y}, \vec{x} \mid \vec{l}) \overset{\Delta}{=} P(\vec{Y} = \vec{y}, \vec{X} = \vec{x} \mid \vec{l}) = m_{i_1} b_{i_1 y_1} a_{i_1 i_2} \cdots a_{i_{N-1} i_N} b_{i_N y_N}\text{，}$$

$$\ln P(\vec{y}, \vec{x} \mid \vec{l}) = \ln m_{i_1} + \sum_{n=1}^{N} \ln b_{i_n y_n} + \sum_{n=1}^{N-1} \ln a_{i_n i_{n+1}}\text{．}$$

由此得

$$Q(\vec{l}^* \mid \vec{l}) = \sum_{\vec{x}} P(\vec{Y} = \vec{y}, \vec{X} = \vec{x} \mid \vec{l})\,(\ln m_{i_1}{}^* + \sum_{n=1}^{N} \ln b_{i_n y_n}{}^* + \sum_{n=1}^{N-1} \ln a_{i_n i_{n+1}}{}^*)$$

$$= \sum_i \ln m_i P(\vec{Y} = \vec{y}, X_1 = i \mid \vec{l}) + \sum_i \sum_{l=1}^{M} \sum_{1 \le n \le N} \ln b_{il}{}^* P(\vec{Y} = \vec{y}, X_n = i) I_{\{l\}}(y_n)$$

$$+ \sum_{i,j} \sum_{n=1}^{N-1} \ln a_{ij}{}^* P(\vec{Y} = \vec{y}, X_n = i, X_{n+1} = j \mid \vec{l}) \overset{\Delta}{=} Q_{\vec{m}}(\vec{m}^* \mid \vec{l}) + \sum_i Q_{\vec{b}_i}(\vec{b}_i^* \mid \vec{l}) + \sum_i Q_{\vec{a}_i}(\vec{a}_i^* \mid \vec{l})\text{，}$$

此处 带 * 的参数都表示修改后的参数，而 $\vec{a}_i^*, \vec{b}_i^*$ 分别表示待修改的参数 A, B 的第 i 列列向量，显见 $\sum_i m_i{}^* = 1, \ \sum_j a_{ij}{}^* = 1, \sum_l b_{il}{}^* = 1$．这样可以通过分别最大化上面等式右边各项得到将模型 $\vec{l}_n = \vec{l}$ 更新为 $\vec{l}_{n+1} = \vec{l}^*$ 的修正模型 $\vec{l}_{n+1}$．而其中每一项都可以用引理１０。１５最大化．这样从最大化就得到如下的结论．

Baum－Welch 算法（递推关系）：

$$\hat{m}_i^{(n+1)} = \frac{P(\vec{Y}=\vec{y}, X_1 = i \mid \vec{l_n})}{P(\vec{Y}=\vec{y} \mid \vec{l_n})} = g_1(i),\qquad (10.24)$$

$$\hat{a}_{ij}^{(n+1)} = \frac{\sum_{n=1}^{N-1} P(X_n = i, X_{n+1} = j \mid \vec{Y}=\vec{y}, \vec{l_n})}{\sum_{n=1}^{N-1} P(X_n = i \mid \vec{Y}=\vec{y}, \vec{l_n})} \overset{\text{记为}}{=} \frac{\sum_{n=1}^{N-1} V_n(i,j)}{\sum_{n=1}^{N-1} g_n(i)},\qquad (10.25)$$

$$\hat{b}_{il}^{(n+1)} = \frac{\sum_{n=1}^{N} P(\vec{Y}=\vec{y}, X_n = i \mid \vec{l_n}) I_{\{l\}}(y_n)}{\sum_{n=1}^{N} P(\vec{Y}=\vec{y}, X_n = i \mid \vec{l_n})} \overset{\text{记为}}{=} \frac{\sum_{n=1, y_n = l}^{N} g_n(i)}{\sum_{n=1}^{N} g_n(i)},\qquad (10.26)$$

其中 $g_n(i) = \dfrac{a_n(i) b_n(i)}{\sum_i a_n(i) b_n(i)}$ 如(10．14)与(10．15)，而

$$V_n(i,j) \overset{\Delta}{=} P(X_n = i, X_{n+1} = j \mid \vec{Y}=\vec{y}, \vec{l}) = \frac{a_n(i) a_{ij} b_{jy_{n+1}} b_{n+1}(j)}{\sum_i a_n(i) b_n(i)}.\qquad (10.27)$$

((10.27)的推导需要先算 $P(\vec{Y}=\vec{y}, X_n = i, X_{n+1} = j \mid \vec{l})$，它与计算 $P(\vec{Y}=\vec{y}, X_n = i \mid \vec{l})$ 的方法几乎完全一样)，此时还有

$$g_n(i) = \sum_j V_n(i,j).\qquad (10.28)$$

在上面所述的算法中，初始值 $\vec{l_0}$ 的设置会直接影响到估计的好坏．为此常用的一种方法是，根据先验知识设置一条较长的"标准" 的虚拟状态链 $\{X_1^{(0)}, \cdots, X_N^{(0)}\}$，用４．０段中的频率估计粗估一个 $\vec{l}$，并把它取为 $\vec{l_0}$．

对于隐 Markov 模型，甚至对于第 15 章中的一般 EM 模型，观测链提供了所有的已知信息．但是，与一般的 EM 方法相比，HMM 远为简单，因为如果我们仔细考察 Baum-Welch 算法，就可以发现它并不依赖于状态链的估计．这就是说，隐马氏模型中的参数估计与状态链的估计之间是解耦的．而一般 EM 算法常常都是有耦合的，即在估计参数时需要知道状态的估计，而在估计状态是又需要知道参数的估计．

在向前与向后两个递推公式基础上的 Baum-Welch 算法，给出了估计模型参数的合理方法．虽然在理论上还没有证明其极限就是最大似然估计．

[注 1] 可以对 HMM 条件作一点小的推广： 假定对 $i \in \{1, \cdots, N\}, l \in \{n_1, \cdots, n_M\}$ 有

$$\begin{aligned} &P(Y_n = v_l \mid X_n = i, X_{n-1} = i_{n-1}, Y_{n-1} = n_{l_{n-1}}, \cdots, X_1 = i_1, Y_1 = n_{l_1}) \\ &= P(Y_n = v_l \mid X_n = j, X_{n-1} = i) \overset{\Delta}{=} b_{ij}(l), \end{aligned}\qquad (10.29)$$

$$P(X_{n+1} = j \mid X_n = i, Y_n = n_{l_n}, X_{n-1} = i_{n-1}, \cdots, X_1 = i_1, Y_1 = n_{l_1}) = P(X_{n+1} = j \mid X_n = i) \overset{\text{记为}}{=} a_{ij}\qquad (10.30)$$

与 n 无关．在实际问题中常常会遇到这一种情形．

[注 2] 在 HMM 中，X_n, Y_n 都还可以是多维的向量， 这时的矩阵 A, B 就相当复杂．

［注 3］ 如果模型中的时间参数n 相应地变为地点参数．那么就导致二维的隐随机场模型．特别对于给定的∂一领域系统，就有隐 ∂一Markov 场模型．此时 Baum-Welch 算法就无法进行了．这时的学习问题可以，例如用如下的交错估计方法(这个方法是耦合的，在这点上它与 Baum-Welch 算法不同，我们仍称这个算法为推广的 Viterbi 算法)：

(1) 先在某个$\vec{l}=\vec{l}^{(n)}$固定下，用推广的 Viterbi 算法得到优化的状态轨道；

(2) 再求$\hat{l}$，使$P(\vec{Y}=\vec{y}\mid\vec{X}=\vec{x}^*,\hat{l})=\sup_{l} P(\vec{Y}=\vec{y}\mid\vec{X}=\vec{x},\vec{l})$，其中

$\vec{x}^*=(x_1^*,\cdots,x_n^*)$ 是随n增加时轨道的最佳估计．也可以使用轨道频率.

(3) 令$\vec{l}^{(n+1)}=\hat{l}$，把$\vec{l}^{(n)}$更新为$\vec{l}^{(n+1)}$;.

(4) 如此反复 迭代.

与之联系的有以下一些问题:

问题 1．用推广的 Vertabi 算法频率算法代替 Baum-Welch 算法后 ，解决了隐 Markov 场时，Baum-Welsh 算法:在二维难于递推的困难．但是如何给出递推公式 ？ 能否在某些条件下,得到递推更新中每次修正都往好的方向发展的结论， 例如,使相对熵增加 ？

问题 2．如果问题 1 中的估计结果较好， 那么用一个随机取的$\vec{x}$代替$\vec{x}^*=(x_1^*,\cdots,x_n^*)$，是否还能接受 ？

问题 3 隐 Markov 模型中，如何引进 Bootstrap 式思维．

［注 4］在生物信息论中，还常用隐半 Markov 模型．这时的做法可以是：

(1) 假定一个初始模型$\vec{l}^{(0)}$(包括初分布，停留时间，转移参数)，由观测资料通过 Viterbi 算法估计轨道；

(2) 由此轨道通过计算频率来更新初分布，停留时间，转移参数，得到更新模型$\vec{l}^{(1)}$；

(3) 重复步骤(1)与(2)，直至较为稳定.

在语音识别中，也有人使用隐半 Markov 模型，但是与用隐 Markov 模型相比较，在评估是否得到显著的改进这一点上，还颇多争议。

5 关于隐 Markov 模型的评注

5．1 隐 Markov 模型包容度大有非常宽的应用面

用隐 Markov 模型建模的优点

(1) 隐 Markov 模型是一种简单的数学模型．它的包容性很大，内涵很广，在应用中的弹性相当大，可以很好地利用我们对于被建模的对象所了解的先验知识，因而有很广的适应性.

(2) 隐 Markov 模型是一种不完全数据的统计模型（ 包含了非线性滤波模型），它之所以被广泛采用，在于这种模型既能反映对象的随机性，又能反映对象的潜在结构，便于利用对象的结构与局部联系性质等方面的知识，以及对研究对象的直观的与先验的了解．

(3) 隐 Markov 模型是很少的几个既从物理模型出发，又与数据拟合相直接联系的算法之一.

(4) 隐 Markov 模型虽然也具有某些黑箱的特点，但是与纯黑箱操作的人工神经网络算法相比，有明显的优点.它的参数常常有较为实质的含义，而线性模型，时间序列模型中的参数一般只是作为被拟合了的参数而出现的，缺少较为真实含义.

(5) 隐 Markov 模型有快速而有效的学习算法.

以上诸点使隐 Markov 模型称为随机建模与拟合的重要基石，在应用中有其普适性.

在具体应用隐 Markov 模型建模时，首先要设定 Markov 链的状态集及其规模，即总状态数 L，这也有相当大的弹性．然后，确定相应的观测过程．为了使得观测链(或场)与马氏链(或场)之间有隐 Markov 关系， 我们常常需要对实际提供的观测链(场)加以改造．例如，一个脱机手写体汉字是一张黑白象素图(一个取值 0 或 1 矩阵)，它和刻划笔道的∂一Markov 场之间的依赖关系并不能满足隐 Markov 关系的要求．需要对它进行预加工，也就要先进行 “特征提取”．以尽量简化得使隐 Markov 关系能够成立．

在确定了模型规模以后，对于隐 Markov 模型的处理，又分为两个不同的相位： 学习相位与运转相位. 学习相位是由给定的观测过程的一组样本出发去估计隐 Markov 模型的参数 $\vec{l}$，$=(m, A, B)$，从而完全确定模型; 而运转相位则是用学习相位所确定的模型参数，对给定的某个观测过程的样本，来估计相应隐 Markov 模型的状态, 或计算出它在各个隐 Markov 模型下出现的概率.

语音识别和脱机手写体汉字识别最终需要知道的是给定的语音或脱机手写体汉字样本应出于哪个模型(哪个参数组)，这是一个识别过程, 通常我们可以按照 Bayes 方法（后验概率方法）去决定模型，即选取使得给定样本出现的概率最大的模型作为它的分类归属，这就是识别的结果． 而在 DNA 碱基序列的排序的问题中，考虑的方式则与之不同. 在 DNA 碱基序列中，需要知道的是如何排序，也就是需要知道给定的某个观测过程的样本所对应的隐 Markov 链的状态.

在学习相位中, 我们面临的是一个不完全数据的参数估计问题． 除了 Baum –Welsh 算法以外，还可以进行交错估计．即可先设置一组初值参数, 由此出发配合观测列按 Bayes 方法去估计马氏链的状态列, 再用此状态列配合观测列用最大似然估计去重新估计模型参数，交替地不断重复此二步骤，…，以达到比较稳定的最终结果. 也可以从设置马氏链的一个状态列作为开始, 去估计参数组, 再由此参数组对应的模型去估计状态, 如此交替进行.

在应用隐 Markov 模型于语音识别或手写体 汉字识别时，往往把一个音, 一个词, 一个字对应于一个隐 Markov 模型 $\vec{l}$ ．这时当然也可以把一个标准音或标准字的状态设置为马氏链的初值，再开始上述交替估计的过程.

值得指出的是，利用已知的先验知识, 来得到简单的递推计算公式, 是使得隐 Markov 模型的计算能够化简的关键. 在实际应用中, 参数组的维数常常非常高，这就使得计算无法实际进行. 所以在交替估计的过程中, 我们还可以借助 Gibbs 采样的思想, 一次只变化一个状态分量或一个参数，以简化计算. 也常常使用 Markov 链 Monte Carlo 方法.

对于隐 Markov 模型的状态是否取得合理, 它们是否包含了所研究的问题的必要的信息，观测过程是否包含了认识模型的足够信息等问题，也是使用中必需了解的. 怎样从我们能够得到的信息(通常主要是样本数据)来判断上述问题，在统计上也是至关重要的.

5．3 隐 Markov 模型的更为一般的形式

隐 Markov 模型的观测链$\{Y_n\}$ 还可以是连续型随机变量，其典型情形有正态分布，Gamma 分布$\Gamma(k, l)$，椭圆对称分布(密度具有 $\frac{1}{|\Sigma|^{\frac{1}{2}}} f((\vec{y}-\overleftarrow{m})^T \Sigma^{-1}(\vec{y}-\vec{m}))$形式)，或上述各种分布的混合分布，例如 $l N(m_1, s^{2})+(1-l) N(m_2, s_2{}^{2})$， $l \exp_{m_1}+(1-l) \exp_{m_2}$ 等等．

更一般地，隐 Markov 模型中的一个状态甚至还可以是一个整个的随机过程．

［注］ 隐 Markov 模型在应用方面的参考文献量十分大，而它在理论方面的参考文献还不多，例如关于大数定律的有

Brain G Leroox, Maximum-liklihood estimation for HMM. Stochastic processes and their Appl. 40(1992), 127-143.

关于中心极限定理的有

(1) P. J. Bickel and Y. Ratof, Inference in hidden Markov models I. (local

asymptotic normality in the stationary case) Bernoulli, Vol.2, No.3, 199-228, 1996 .

(2) P. J. Bickel, Y. Ratof and T. Ryden , Asymptotic normality for the maximum likelihood estimates for general hidden Markov models. Annals of Statistics, Vol. 26, No.4, 1614-1635, 1998.

６． 隐 Markov 模型的应用例子梗概

６．１ 语音的机器识别

语音识别与下面６．２段的脱机手写体汉字识别，这两个问题都有一个共同的特点：它们都有随机性，但是也都有一个潜在的基本结构．例如,一个语音也有其基本结构，而发音时口型的大小,长短,强弱，在口腔中的位置等，却因人而会有随机的变化．即使是同一个人，在不同时间发同一个音也是有随机差异的．而一个汉字的基本笔画及相对关系是确定的,但是笔画的长短，间隔与倾斜度却因人而异．用隐 Markov 模型来描述这种具有结构的随机性相当有效．

这里以简单化了的实际问题为例，这样可以使我们摆脱许多细节而突出思路．一个词的语音可由许多音素组成．如前所述，隐 Markov 模型在语音识别中，首先是建立一个对应关系，使一个音对应于一个隐 Markov 模型(或几个隐 Markov 模型)．这里的隐 Markov 模型的状态，一般取为这个音所含的全部可能的音素(或其细分,或其组合)．例如一个汉语 “ 高 ” 的发音可能由

$$‘ ggg(ge)(ga)(ga)(ga)\ aaa(ao)(ao)ooo ’ \qquad (10.31)$$

(后面可加上足够的□(空))等音素组成，而且每个音素究竟维持多少时间单位，对不同的人，对同一人在不同的时刻发这个音，都可能有很大的不同．于是对应于“ 高 ”这个音，是一列相继的状态组成的一个现实(在上面序列中一个字母代表相应汉语拼音，而一个括号中有两个字母则表示过渡)．由于声转移有内在统计规律，我们就可以自然地把它考虑为一个 Markov 链 (X_n) 的状态序列．但是，(X_n) 实际上并不可观测．即我们能测量到的并非这个音对应的 Markov 链本身，而是每个字母对应的声波信号的振幅的采样(或经过某种预处理的一些特征量)．这个可以观测到的现实就是对于一个音的观测链样本列．状态链与观测链在一起就组成了一个隐 Marlov 模型，其中的状态序列，就是该隐马氏模型的马氏链的轨道(包括 ”□（空）”)．所以说一个音对应于一个隐 Markov 模型(由于发音人音质的差异，一个音也可以对应与多个隐 Markov 模型)．一般在语音识别中，把音频信号直接作为观测值常会使计算时的数据量过大．更为实际的是把音频信号的一些特征量，作为压缩了的观测数据，这称为预处理．这些特征可以是音频信号的过零率，线性预测系数(LPC，即 Linear Predict Coefficient，它是把采样值作为 AR 模型（参见第１１章）拟合后，得到的模型的系数)，倒频系数(CEP: 记信号 $s(n)$ 的离散 Fourier 系数为 $F(s(n))$，则 $CEP \overset{\Delta}{=} F^{-1}(\ln|F(s(n))|)$)等等．一般地常取 12 至 16 个 LPC 作为压缩了的观测数据．也就是把原始观测数据用 AR(p) 模型 (其中 p = 12 至 16)建模．再把这样拟合的参数作为压缩了的观测数据．(为了保证状态到观测的转移只依赖于前一个状态，有时还需要把邻近的状态 “ 堆 ” 起来，例如说，把相继出现的 k 个状态看成为一个状态向量，并把与此 k 个状态向后错一位的相继出现的 k 个状态，看成为其后时刻的一个状态向量，…，使一个时刻的 “ 状态 ” 或 “ 观测 ” 成为一个向量)．再则，有时也常把在状态已知的条件下，观测向量的条件分布假定为(一维的或多维的)正态分布(一般假定方差矩阵为对角矩阵)．总之，我们可以把一些特征量作为观测链 $\{Y_n: n \geq 1\}$，其中 Y_n 可以是多维的(向量)．序列(10.31)并不能观测到．它正是我们想要知道的状态链．为了解它，就要

由对该音的一组观测样本(若干个发这一音的声信号或提取的一些特征量)，去作相应的隐Markov 模型的参数 $\vec{l} = (m_0, A, B)$ 估计，这就是学习相位．于是语音的识别对应于设置一些模板：一个音对应于一个或数个模板 $\vec{l}$ （有时也可以用最佳状态列的估计作为音或词的代表)．这时模板的代表性，选多少个状态，选多少个观测，对于计算量的控制与，都至关重要.为了能有效地学习，就需要有足够多的样本，也就是要有这个音或词的重复发音．例如把 m 个不同的人的对同一个音或词的发音的预处理加工，作为 m 条观测链样本．学习的结果是建立了标准人的发音与隐 Markov 模型间的对应(也可以让一个音或一个词对应于几个隐马氏模型)，以得到一个模板库．

在学习得到模板后, 就可以作语音的机器识别问题，即对于一条新给的观测链, 在一族(可能是上千个)已建立好的隐 Markov 模型的模板中, 选取一个与之“ 最近的 ”(相对熵最小的) 隐马氏链模型 $\vec{l}$，它所代表的音或词, 就可以判定为这个观测所对应的音或词, 这就是识别问题.

隐 Markov 模型的基本假定是，状态链是时齐的 Markov 链, 而且状态到观测的转移也是时齐的. 但是, 在实际语音识别中，常会遇到非时齐的情形. 这时如果照搬时齐情形的做法, 就会引起振荡． 然而, 用非时齐的隐马氏模型来建模, 无论在理论上或对实际计算, 都远非可取. 所以运用隐 Markov 模型时极需要灵活处理. 实际上, 在实践中遇到的问题, 绝非只是理论的直接应用, 其复杂的程度常常不比理论问题差．

如果状态链在一个状态上的停留时间分布，显著地不是指数分布(离散采样时它对应于几何分布)，那么在理论上可以想象，应该改用隐半马氏模型代替隐马氏模型．可惜的是，这种修改后的模型的计算远为复杂．而在实际使用中，这两个模型的差别也并不如我们想象的那样大．因之这种修改也并没有被广泛采用．

6．2　脱机手写体汉字识别

汉字的模板称为码本．在线手写体汉字或数字可以看成一维的，用隐 Markov 模型的框架建模就较为简单．而脱机手写体汉字是二维的，即在隐 Markov 模型中的时间参数 n 变成为二维的空间参数 (n, m)．而二维的 Markov 性就远较一维复杂，因为它没有很自然的序．在确定相应的观测过程时,为了使得观测链(场)与状态链(Markov 场)之间有象(10.3)及(10.4)那样的简单关系，我们常常需要对实际提供的观测链(场)加以改造．例如，在实际上我们得到的手写体汉字是一张黑白象素图(一个取值 0 或 1 矩阵),它和笔画的隐 Markov 场之间的依赖关系不能满足(10.3)或(10.4)的要求．我们要对它进行加工,也就是“特征提取”.尽量简化得使(10.3)或(10.4)能够成立．为了使脱机手写体汉字纳入隐 Markov 场的框架，选取什么作为状态与观测才能较好地体现 Markov 性，这是具体操作中的最重要的．粗略地说，一般可以选取一个格点上八个方向的笔道作为观测的特征．

此外， 数字识别，文字识别， 印刷中的错别字自动检测， 航海中的动态手势识别等等, 也可以用隐 Markov 模型处理，只是有的情形的"时间参数" n 应该相应地改为空间参数．

6．3　DNA 序列片断装配及启动子识别

在遗传学中，生物体属性的遗传是由细胞核中的染色体中所含的基因所指导．基因是由DNA (脱氧核糖核酸) 碱基的序列组成．如果把基因看成语句，DNA 的 A,T,G,C 4 种碱基，就象语句中的字母．一条 DNA 链就是一个 A,T,G,C 这 4 个字母组成的序列．读出 DNA 碱基序列对于基因工程，遗传病及许多癌症的发现与治疗都具有重要意义． 然而，人类基因组计划目的并不是解决“人类由甚么组成的”，而是要回答 “是甚么使得一种有机体区别于另一种有机体”．一个人的基因组的全序列，是一个包含约 30 亿个由 A,T,C,G 组成的序列. 它们分为 23 对染色体，与一条性染色体．生物体中最小的 DNA 链, 约含一百万碱基对(bp, base pair)，即由 A,T,C,G 4 个字母组成的长度为一百万的一个序列． 基因是染色体 DNA序列的一些子序列, 它们给出了有关蛋白质的信息. 至今, 人们知道在基因之间还存在 DNA 序列, 可是还不清楚它们的作用． 一个人的基因组的全序列可以说是他的生命的设计书. 它能为了解遗传与发展等生命

奥秘奠定基础. 而得到基因组的全序列，只是我们读懂这本生命之书的起始的一步. 生物信息论的第一步是 DNA 序列的分析，包括 DNA 序列的比对(alignment)，拼接(assembly)；寻找基因(gene finding)，调控研究(regulation), 基因功能(function)研究等.

(1999 年 12 月 1 日, 人类基因组科学家完成了人类 22 号染色体的测序, 这是在 23 对染色体中第二小的染色体, 有 3350 万个碱基对, 由长短不同的 12 个 DNA 片断组成. 最长的片断有 2300 万碱基对, 其它的片断, 长度在 1000 碱基对到 58.3 碱基对之间. 在 22 染色体上发现了 679 个基因, 其中 545 个是功能基因, 与先天心脏病, 低免疫性, 精神分裂, 智力低, 许多恶性肿瘤如白血病等有关. 另外的 134 个不是功能基因. 2000 年 5 月 8 日, 人类基因组又基本上给出了第 21 号染色体的基因图谱. 这是最小的染色体, 有 225 个基因, 与白血病, 痴呆, 肌肉萎缩, 燥狂性抑郁, 部分癌症等有关. 人类基因的总数比水稻要少, 说明人类的基因可能更为有效).

然而, DNA 序列是非常长的链，用目前的设备是无法读出的. 为了读取一条链，就需要线把它打断成一段一段, 使每一段都可以用设备读出. 而在解读 DNA 序列时, 遇到的困难是：

(1)目前快速读出 DNA 序列的设备只能读出几百到一千个碱基对的链.

(2)快速打断 DNA 序列的方法是随机的, 不能控制断点的位置及段数.

(3)打断后的片断混为一体无法区分次序先后.

解决困难的一个办法是，同时对同一个 DNA 序列的多个拷贝(copy) 作随机打断及作随机片断读出, 利用断点几乎概率为 1 地不会在同样位置重复, 并利用不同的拷贝片断的相互核对, 将片断连接起来(Assembly)得到整体的排序. 但是, 在读出时可能有错读,漏读及误增字符等错误. 且在原序列中还有重复的片断，这就使拼接问题十分复杂. 上世纪末的人类基因测序中，人类基因组的科学家对染色体 DNA 序列，利用一些只出现一次的序列(称为 marker)，将整条序列分段(段长为 2M - 3M)读出并拼接起来，以降低拼接的复杂性. 而 Cerela 公司的科学家则更着重于技术与算法，他们利用最新的毛细管测序机，可以从一段序列的两头同时测得两个 500-700bp 小段 DNA 序列的这个特点，将人类基因组科学家拼接思路中忽视的匹配(mate)信息利用起来，为不用 marker 直接测序开辟了新路.

- 读出了全部基因组的 DNA 序列后，基因检测是后基因处理中的首要问题. 即要知道序列中各部分的含义(如: 哪里是基因，哪里是基因的蛋白编码部分,哪里是其调控部分,…). 不知道这些，DNA 测序的结果不会有用. 每个基因有几部分构成：启动子(Promoter) – 像是基因的引言；外显子(Exons) – 编码部分，它在基因 DNA 序列中也可能是被内含子(introns)分隔的子序列；内含子(Introns) – 外显子间的序列，它的功能至今还不清楚；终止子(Terminator). 它们可示意为:

启动子　外显子 1　内含子 1　外显子 2　内含子 2　外显子 3　终止子

对于已读出的 DNA 序列，Karlin 和 Berger 有一个基因检测的软件，称为 Genescan.利用基因编码区的统计特征，即在没有漏读(也称删除, 记为 D)与没有增读(也称插入, 记为 I)时, 利用 DNA 编码序列具有很强的 3-连子特性(A,T,C,G 4 个核酸组成 64 个 3 – 连子, 即 3 个相连的字母组成的一个编码. 在此６４个编码中除了 3 个编码代表终止符号以外, 其它的６１个编码分别代表 20 个氨基酸 (amino) 的编码(一个氨基酸可以有几个不同的编码), 在编码区无漏读或无增读的条件下，Karlin 和 Berger 采用了动态规划方法，编制了基因检测软件 Genescan，达到了很高的识别率. 但是在编码区有漏读或有增读时，就会发生 3-连子的错误，且随后的所有 3-连子就全被搞乱，接着就会出现一系列的错误，如果还用 Genescan 就会发生严重的错误. 所以，在用这个软件前，必需先在观测序列中，标识错读，漏读(删除)或增读(插入)的碱基对，以便避免出现大量的编码错误，并尽可能的校正错读的碱基对.

对于标识错读，漏读或增读碱基对，G.Churchill 首次提出了用隐 Markov 模型解决 DNA 片断装配问题. 他的基本思想是将由随机枪打断长链中的每个碱基对，附以读出的属性，即读对 R，错读 M，漏读 D 或增读 I (其中最误事的是漏读 D 或增读 I，因为它们引起一系列后继错误)，把这些属性看成一个 Markov 链的状态的样本，而把读出的碱基序列片段看成为相应于这个状态链的观测链. 即设观测过程为可能带有小概率的漏读 D 与增读 I （在实际解读中，漏读与增读度不可能很多）的 DNA 码区序列. 这时，隐 Markov 链的状态空间就是 $\{R, M, D, I\}$，其中 R 表示读对的碱基对, M 表示读错的碱基对. 而隐马氏链的样本列, 就是对观测序列上的

碱基对标出的属性R,D,M 与I 的序列. 这样的考虑是，基于在编码区中的序列上, 氨基酸编码的转移则相当集中. 因而, 隐 Markov 模型能很好地刻画与鉴别有无删除, 错读与插入（增读）, 这样就把问题纳入了隐 Markov 链的框架. 再利用统计性质，就可以纠正在使用寻找基因的软件 Genescan 时可能导致的错误. 在学习相位(即学习问题)，由给定的观测过程的一组样本找出隐 Markov 模型的参数族，可以用 Baum-Welch 算法. 而在运转相位(即解码问题)，对给定的观测样本找出相应的 Markov 模型的状态，可以用 Viterbi 算法. (相仿的思路也出自借鉴 Gibbs 场的思想，即在学习时先设定初值，再轮番修改参数和状态 (见第 15 章中的 EM 算法). 而在运转时用 Bayes 决策找出最佳状态).

在 DNA 序列中，启动子是一个特殊位点，它反映了在 DNA 转录过程中发挥定位功能的 DNA 片断表现在序列上的规律性. 所以, 启动子识别是生物信息论中首先要解决的问题. 隐 Markov 模型也可以用来识别启动子(Promotor)等特殊位点及编码区. (这些特殊位点还包括增强子(enhancer)，抑制子(silencer))，转录起始位置(TSS, transcription starting site)，外显子(exon)，内含子(intron)，受子(accepter) ,给子(donor)等等).

利用已有的生物学知识及数据库，构造合适的数学模型，例如,合适的隐 Markov 模型，再与数据进行反复的交叉验证与修改，是得到成功的识别方法的关键.

习题 10

1. 证明 在均值为 m，方差为 s^2 的分布密度中，正态分布的熵最大； 而在均值向量为 $\vec{m}$，方差矩阵为 Σ 的多维密度中，Gauss 分布 $N(\vec{m},\Sigma)$ 的熵最大. 请读者自己写出最大熵的表达式.

2. 证明在有限区间$[a,b]$ 上取值的分布密度中, 均匀分布的熵最大. 这个熵为 $\log(b-a)$. 再把这个结论推广到多维情形.

3. 设 $g_1(x),\cdots,g_m(x)$ 都是非负有界函数. 那么, 在满足条件

$$\int g_i(x)p(x)dx = a_i, (i \le m)$$

的分布密度 $p(x)$ 中, 当且只当

$$p_0(x) = Ce^{-(l_1 g_1(x)+\cdots+l_m g_m(x))}$$

时熵为最大, 其中

$$C = \frac{1}{\int e^{-(l_1 g_1(x)+\cdots+l_m g_m(x))}dx},$$

而 $l_1,\cdots,l_m$ 是由

$$\int g_i(x)e^{-(l_1 g_1(x)+\cdots+l_m g_m(x))}dx = a_i\int e^{-(l_1 g_1(x)+\cdots+l_m g_m(x))}dx, (i \le m)$$

确定的正数.

4. 能不能把 3 推广到 $g_1(x),\cdots,g_m(x)$ 不全为非负, 也不全为有界的情形 ?

龚光鲁，钱敏平著　应用随机过程教程 – 与在算法和智能计算中的应用
清华大学出版社，2003

第 11 章 Gauss 系, 二阶矩过程 时间序列

1. 全体方差有限的随机变量构成的 Hilbert 空间

1, 1　实值情形

期望为 0 且方差有限的随机变量全体构成的集合, 记为 L^2．它是一个无穷维的线性空间．在 L^2 上可以仿照欧氏空间定义内积:

$$(x,h)\overset{\Delta}{=}E(xh),\ (\forall x,h\in L^2).$$

于是 L^2 在此内积下成为(无穷维）欧氏空间． 再仿照欧氏空间定义模（相当于长度，也称为均方距离)

$$\|x\|\overset{\Delta}{=}(x,x)^{\frac{1}{2}}.$$

它有以下的重要性质:

（1）Cauchy 不等式: $|(x,h)|\le\|x\|\cdot\|h\|$.

（2）三角形不等式: $\|x+h\|\le\|x\|+\|h\|$.

这样 $d(x,h)\overset{\Delta}{=}\|x-h\|$就是定义在 L^2 上的一个距离，从而有了收敛性：即若 $\|x_n-x\|\to 0$，则称 $x_n\to x$ (它的含义恰是均方收敛 $E|x_n-x|^2\to 0$).作为线性空间 L^2 除了不再是有限维以外，它与普通的有限维欧氏空间一样，作为距离空间都是完备的，即：随机序列$\{x_n\}$在 L^2 中收敛的充要条件为它是一个 Cauchy 列(见第 9 章)，也就是

$$\forall e>0,\exists N,\text{只要}n,m>N,\text{就有}\|x_n-x_m\|<e.$$

定义１１.１　完备的无穷维欧氏空间常称为 Hilbert 空间．

1．2　复值情形

在通信等领域，人们常用复值的量．记 i 为虚数单位，若 x， h 都是随机变量，则 $V=x+ih$ 称为复值随机变量，它具有复的数学期望 $EV=Ex+iEh$，非负的方差

$$\operatorname{var}V\overset{\Delta}{=}E(V\overline{V}\}=Ex^2+Eh^2.$$

用一点技术处理可以证明(本书从略)：对于任意复常数 a，恒有 $E(aV)=aEV$．两个期望为 0 的复随机变量V_1,V_2的内积定义为 $(V_1,V_2)\overset{\Delta}{=}E(V_1\overline{V_2})$．于是, 实值情形的结论对于复值情形也是适用的.

2　随机变量族的均方信息空间与滤波

2．1　均方信息空间

记号１１．２　由随机变量族$\{x_a:a\in I\}$中任意有限个元素的任意有界连续函数全体组成 L^2 的一个线性子空间(但对极限并不封闭)，记为 $\Phi(x)$．

再记$\overline{\Phi}(x)$为：包含 $\Phi(x)$ 且对 L^2 中收敛性封闭的最小集合，那么 $\overline{\Phi}(x)$ 是 L^2 的 Hilbert 子空间．我们称它为$\{x_a:a\in I\}$的均方信息空间．这里 ”均方信息” 的含义是指只要 “ 方差有限 ”的信息.

2.2 滤波问题

滤波问题的一般提法为：假定随机变量族$\{x_a : a\in I\}$是可以实际测量的，且期望为 0．而另一个随机变量h 的期望为 0，方差有限，但是不能实际测量得到．我们需要从测得的$\{x_a : a\in I\}$对随机变量h 给出一个估计$\hat{h}$． 这样的问题称为滤波问题．

解滤波问题的思路的一般模式

我们知道在欧氏空间中，对于一个子空间外一个点，求此点在该子空间上的投影，就是在该子空间中寻找一个与它距离最近的点．这个事实在 Hilbert 空间仍然正确（由于 Hilbert 空间是无限维的，欧氏空间情形的证明当然不再适用．我们略去 Hilbert 空间中的这个基本事实的证明)．于是滤波问题的解$\check{h}$ 就是h在 L^2? Hilbert 子空间$\overline{\Phi}(x)$上的投影．即下面的命题

命题１１.３ 在随机变量族$\{x_a : a\in I\}$的均方信息空间$\overline{\Phi}(x)$中的元素$\check{h}$与随机变量h 最近，即$\| h-\check{h}\| = \min_{V\in\overline{\Phi}(x)} \| h-V\|$ ，的充要条件是

$$\check{h}\in \overline{\Phi}(x)\text{，且 } h-\check{h}\perp \overline{\Phi}(x),$$

即对任意n 及任意n元（Borel）函数 $g^{(n)}$，恒有

$$E[(h-\check{h})g^{(n)}(x_{a_1},\cdots,x_{a_n})]=0\,. \qquad (11.1)$$

定义１１.４ 将$\check{h}$ 记为$\mathrm{Proj}_{\overline{\Phi}(x)}h$ ，称为h 关于$\{x_a : a\in I\}$的(非线性)滤波．

按条件期望的定义，(11. 1)说明了

$$\mathrm{Proj}_{\overline{\Phi}(x)}h = E(h\,|\,x_a : a\in I)\,, \qquad (11.2)$$

即h 关于$\{x_a : a\in I\}$的条件期望．于是在特殊情形，即$\{x_a : a\in I\}$只有一个随机变量x 的情形， 我们有

(1) 若(x,h)是离散的随机变量，$P((x,h)=(x_i,y_j))=p_{ij}$时，$\check{h}=\sum_i x_i\dfrac{p_{ih}}{\sum_i p_{ih}}$．

(2) 若(x,h)有密度 $p(x,y)$时，$\check{h}=\int x[\dfrac{p(x,h)}{\int p(u,h)du}]dx$．

3 Gauss 系与投影再访

3.1 Gauss 过程的定义与等价条件及其性质

复习１１.５ 随机向量$\vec{h}=\begin{pmatrix}h_1\\ \vdots\\ h_n\end{pmatrix}$称为服从$n$维 Gauss 分布，记为$\vec{h}\sim N(\vec{m},S)$，

如果其特征函数有以下形式

$$Ee^{i\vec{l}^T\vec{h}}=e^{i\vec{l}^T\vec{m}-\frac{1}{2}\vec{l}^T\Sigma\vec{l}}\,,$$

其中$\vec{m}$ 是n 维向量，Σ 是$n\times n$非负定矩阵（即任意n 维向量$\vec{x}$，恒有$\vec{x}^T\Sigma\vec{x}\geq 0$）．显见有：

随机变量$\vec{h}=\begin{pmatrix}h_1\\ \vdots\\ h_n\end{pmatrix}$服从$n$维 Gauss 分布，当且仅当，对于任意实数 $a_1,\cdots,a_n$ ，线性组合

$\sum_{k=1}^{n}a_k h_k$ 服从一维 Gauss 分布（ 即或者是一维正态分布, 或者是常数）.

请读者验证.

命题１１．６ n维随机向量$\vec{h}\sim N(\vec{m},S)$，当且仅当， 存在m和相互独立的m个服从$N(0,1)$的随机变量$Z_1,Z_2,\cdots,Z_m$，以及$n\times m$矩阵 $A=(a_{ij})$ $(i\le n, j\le m)$，使

$$\vec{h}=A\vec{Z}+\vec{m}\text{，}\Sigma=AA^T\text{，}$$

其中$\vec{Z}=\begin{pmatrix}Z_1\\ \vdots\\ Z_m\end{pmatrix}$. 矩阵$\Sigma$非退化（行列式不为 0 时）的 Gauss 分布，称为多维正态分布.

命题１１.７ 若$\vec{h}\sim N(\vec{m},S)$，则有

(1) $E\vec{h}=\vec{m}, E[(\vec{h}-\vec{m})(\vec{h}-\vec{m})]=\Sigma.$

(2) $A\vec{h}+\vec{b}\sim N(A\vec{m}+\vec{b},A\Sigma A^T)$.

(3) $\vec{x}^{(n)}$服从 Gauss 分布, 且$\vec{x}^{(n)}\xrightarrow{p}\vec{x}$(指所有分量都概率收敛), 则$\vec{x}$服从 Gauss 分布.

请读者验证(1)与(2)). 对于(3)，我们只证一维情形，因为多维情形类似. 由假定$x^{(n)}\xrightarrow{p}x$，对于ε >0，有

$$P(|x^{(n)}-x^{(m)}|>e)\to 0, n,m\to\infty\text{ .}\qquad (11.3)$$

记 $m_{ij}=Ex^{(i)}-Ex^{(j)}, s_{ij}^{\ 2}=Var(x^{(i)}-x^{(j)})$，$\Phi(x)=\frac{1}{\sqrt{2p}}\int_{-\infty}^{x}e^{-\frac{u^2}{2}}du$. 于是(11.3)变成在$n,m\to\infty$时

$$\frac{1}{\sqrt{2p}\,s_{ij}}\int_{|x|>e}e^{-\frac{(x-m_{ij})^2}{2s_{ij}^{\ 2}}}dx\to 0\text{ .}$$

而上式左边等于

$$\frac{1}{\sqrt{2p}}\int_{|s_{ij}y+m_{ij}|>e}e^{-\frac{y^2}{2}}dy=[1-\Phi(\frac{-m_{ij}+e}{s_{ij}})]+\Phi(-\frac{m_{ij}+e}{s_{ij}})\text{ .}$$

因此(11.3)蕴含：存在N，只要$i,j>N$，上面右式就小于$1-\Phi(1)(=\Phi(-1))$. 由此

$$[1-\Phi(\frac{-m_{ij}+e}{s_{ij}})]<1-\Phi(1)\text{，}\Phi(-\frac{m_{ij}+e}{s_{ij}})<\Phi(-1)\text{ .}$$

从而在$i,j>N$时有

$$\frac{-m_{ij}+e}{s_{ij}}>1\text{，}\ \frac{m_{ij}+e}{s_{ij}}>1\text{ .}$$

即 $s_{ij} \pm m_{ij} < e$，把此两式相加就得到 $s_{ij} < \frac{e}{2}$（$i, j > N$）. 由此进一步得到 $\pm m_{ij} < \frac{e}{2}$，即 $|m_{ij}| < \frac{e}{2}$（$i, j > N$）. 可见

$$E(x^{(i)} - x^{(j)})^2 = s_{ij}^{\ 2} + m_{ij}^{\ 2} \to 0, (i, j \to \infty).$$

于是 $\{x^{(n)}\}$ 是 Hilbert 空间 L^2 中的 Cauchy 列. 用完备性就得到：$E|x^{(n)} - x|^2 \to 0$. 从而有 $Ex^{(n)} \to Ex, Var(x^{(n)}) \to Varx$，由此

$$Ee^{ilx} = \lim_{n\to\infty} Ee^{ilx^{(n)}} = \lim_{n\to\infty} e^{ilEx^{(n)} - \frac{1}{2}l^2 Var(x^{9n)})} = e^{ilEx - \frac{1}{2}l^2 Varx}$$

可见 x 也服从 Gauss 分布.

定义１１．８ 随机变量族 $\{x_t : t \in I\}$ 称为 Gauss 系, 如果 $\forall n, \forall t_1, \cdots t_n \in I$, $(x_{t_1}, \cdots, x_{t_n})$ 服从 Gauss 分布. 如果 Gauss 系中的指标集 $I = [0, \infty)$, 则称为 Gauss 过程.

对于 Gauss 过程而言，其期望函数 $m(t) \overset{\Delta}{=} Ex_t$ 及相关函数 $B(s,t) \overset{\Delta}{=} E(x_s x_t)$ 完全地确定了它的有限维分布族：

$$\begin{pmatrix} x_{t_1} \\ \vdots \\ x_{t_n} \end{pmatrix} \sim N\left(\begin{pmatrix} m(t_1) \\ \vdots \\ M(t_n) \end{pmatrix}, \begin{pmatrix} & & \\ & \Sigma(t_i, t_j) & \\ & & \end{pmatrix}_{i,j \le n} \right),$$

其中 $\Sigma(t_i, t_j) = B(t_i, t_j) - m(t_i)m(t_j)$, 称为协方差函数.

例１１．９

(1) 若 (x, h) 为正态分布, 则 $(x, x-h, x+h)$ 为 Gauss 分布

(2) 若 $\{x_n : n \ge 1\}$ 为 Gauss 系, $h_n = \sum_{k=1}^{n} a_{nk} x_k$, 则 $\{h_n : n \ge 1\}$ 为 Gauss 系.

记号１１．１０ 期望为 0 的全体 Gauss 随机变量不仅是 L^2 的线性子空间，而且由于均方收敛蕴含概率收敛(用 Chebyshev 不等式)，由命题１１．７的（３）可知，它还是 Hilbert 子空间．记

$$L(x) \overset{\Delta}{=} \{x_t : t \in I\} \text{中任意有限个元素的实线性组合组成的集合};$$

称为 $\{x_t : t \in I\}$ 的线性包. 再记 $\overline{L}(x)$ 为包含 $L(x)$, 且对于均方极限封闭的最小集合: 即

$$\text{若} h^{(n)} \in \overline{L}(x), \| h^{(n)} - h \| \to 0, \text{则} h \in \overline{L}(x).$$

$\overline{L}(x)$ 称为 $\{x_t : t \in I\}$ 的线性闭包。它的每个元素都是 $\{x_t : t \in I\}$ 中元素线性组合在概率意义下的极限，因而可看成整个随机过程 $\{x_t : t \in I\}$ 的某个"线性泛函" $\Phi(x)$．作为命题 11.7 的直接推论，我们有下述命题

命题１１．１１(封闭性命题) 如果 $\{x_t : t \in I\}$ 是 Gauss 系，则 $\overline{L}(x)$ 也是 Gauss 系.

此外，我们还有

命题１１.１２（独立性命题)

(1) 若 $\{x_a, h_b : a \in I, b \in J\}$ 是 Gauss 系, 则 $\{x_a : a \in I\}$ 与 $\{h_b : b \in I\}$ 独立的充要条件为: 对于任意 x_a, h_b 都有 $Cov(x_a h_b) = 0$.

(2) 若$\{x_a: a\in I\}$与$\{h_b: b\in J\}$独立，那么$\overline{\Phi}(x)$与$\overline{\Phi}(x)$独立.

3.2 Gauss 过程的投影 —线性滤波

定义11.13 设$\{x_t: t\in I\}\bigcup\{h\}$是期望为 0 的 Gauss 系．随机变量 h 在 Hilbert 子空间$\overline{L}(x)$上的投影，记为$\hat{h}$，称为h关于$\{x_a: a\in I\}$的线性滤波．$\hat{h}$也记为$\mathrm{Pr}\,oj_{\overline{L}(x)}h$，即

$$\hat{h}\in\overline{L}(x),\ 且\ h-\hat{h}\perp\overline{L}(x). \qquad (11,1)'$$

(11.1)’称为线性投影公式．

定理11.14 设$\{x_t: t\in I\}\bigcup\{h\}$是期望为 0 的 Gauss 系．那么

$$\check{h}=\hat{h},$$

即：对 Gauss 系而言，非线性滤波与线性滤波是一样的．

证明 首先注意$\hat{h}\in\overline{L}(x)\subset\overline{\Phi}(x)$．其次，由于$h-\hat{h}\perp\overline{L}(x)$，我们有$E[(h-\hat{h})x_t]=0,(\forall t\in I)$．由命题11.12得到，$h-\hat{h}$与$\{x_t: t\in I\}$独立，因而也与$\overline{\Phi}(x)$独立．于是对于任意对$V\in\overline{\Phi}(x)$有 $E[(h-\hat{h})V]=0$．这正说明了$h-\hat{h}\perp\overline{\Phi}(x)$．从而$\check{h}=\hat{h}$．

3.3 复 Gauss 过程

设$V_t^{(k)}=x_t^{(k)}+ih_t^{(k)},(k=1,2)$的期望为 0，则二元函数$B(s,t)\overset{\Delta}{=}E(V_s^{(1)}\overline{V_t^{(2)}})$称为过程$\{V_t^{(1)}\}$与$\{V_t^{(2)}\}$的相关函数．它是一个复的非负定函数，即对于任意$m$，$t_1,\cdots,t_m$及任意复数$a_1,\cdots,a_m$，恒有

$$\sum_{k,l=1}^{m}B(t_k,t_l)a_k\overline{a_l}\geq 0.$$

定义11.15 复随机过程$V_t=x_t+ih_t$称为复 Gauss 过程，如果$\{x_t\}$与$\{h_t\}$是相互独立，且有限维分布族相同的 Gauss 过程．

3.4 Gauss 过程的特征泛函

对于期望函数为 0，协方差函数为$R(s,t)$的 Gauss 过程x_t及任意连续增函数$F(t)$，定义 Gauss 过程x_t的特征泛函为

$$\overline{\Phi}_x(F)\overset{\Delta}{=}Ee^{i\int_0^T x_t dF(t)},$$

即它是 Gauss 随机变量$\int_0^T x_t dF(t)$的特征函数在 1 处值．由于$E[\int_0^T x_t dF(t)]=0$，$Var[\int_0^T x_t dF(t)]=\int_0^T\int_0^T R(s,t)dF(s)dF(t)$，因此

$$\overline{\Phi}_x(F)=e^{-\frac{1}{2}\int_0^T\int_0^T R(s,t)dF(s)dF(t)}.$$

4. 平稳性与宽平稳性

4．1　平稳序列与宽平稳序列.

在实用领域中称随机变量序列$x_n(-\infty<n<\infty)$为时间序列， 常假定其方差有限.

定义１１.１６　如果对于任意m,k，m维随机变量$(x_{k+1},\cdots,x_{k+m}\}$都与$(x_1,\cdots,x_m\}$同分布， 则称$x_n(-\infty<n<\infty)$为平稳序列.

如果有对于任意n,k有

$$Ex_n=\text{常数}m,\qquad E(x_nx_{n+k})=\text{某个不依赖}n\text{的函数}R(k),$$

则称$x_n(-\infty<n<\infty)$为宽平稳序列, 而$R(k)$称为它的相关函数.

宽平稳序列是在应用中最常用的时间序列．它代表数学期望函数$m(t)\overset{\Delta}{=}Ex_t$与相关函数$B(t,t+k)\overset{\Delta}{=}E(x_tx_{t+k})=R(k)$这两个最重要的平均特征都不依赖时间$t$的时间序列. 通常假定宽平稳序列的期望函数为零, 否则可以预先减去它的期望. 在理论上宽平稳列常在L^2的框架中讨论.

定义１１.１７　在工程, 无线电, 控制等诸多领域中出现的宽平稳序列常假定其期望为0, 而其相关函数$R(k)$常可表成某个非负函数$f(l)$的 Fourier 系数:

$$R(k)=\int_0^{2p}f(l)e^{ikl}dl\,,\tag{11.4}$$

这个非负函数$f(l)$称为此宽平稳序列的谱密度.

相关函数概括了宽平稳序列的最重要的统计特征, 因而谱密度$f(l)$也同样概括了宽平稳序列的最重要统计特征. 在实际问题中, 可以通过序列x_n的一段观测值来估计, 即用

$$\hat{f}(l)=\frac{1}{n+1}|\sum_{k=0}^{n}x_ke^{-ikl}|^2\tag{11.5}$$

来估计谱密度$f(l)$, 称为谱图估计. 这个估计简单易用, 但是较为粗略. 参照非参数统计中的核估计的思想, 可以得到一些改进的估计. 而在宽平稳序列的期望不等于0时, 则用

$$\hat{\hat{f}}(l)=\frac{1}{n+1}|\sum_{k=0}^{n}(x_k-\bar{x})e^{-ikl}|^2,(\bar{x}=\frac{1}{n+1}\sum_{i=0}^{n}x_i)\qquad(11.4)'$$

作为谱图估计.

在应用中，可以认为相关函数与谱密度有相同的作用．但是在具体处理上，又各有其长处．用相关函数研究宽平稳序列称为时域方法，而用谱密度研究宽平稳序列则称为频率域方法. 具有谱密度的宽平稳序列可以用它的样本平均来估计其均值m:

$$\|\frac{x_0+\cdots+x_n}{n+1}-m\|\to 0\,,$$

此公式的收敛是 L^2? 的收敛,就是均方收敛．其证明需要用宽平稳列的谱积分理论，这是一个已经发展得过于成熟的理论，本书由于篇幅的限制而不再选入这部分材料．又因为均方收敛蕴含了概率收敛，所以有： 以接近于1的概率有$\frac{x_0+\cdots+x_n}{n+1}\approx m$. 同时还可以用样本来估计相关函数：

$$\|\frac{x_0x_k+\cdots+x_nx_{n+k}}{n+1}-R(k)\|\xrightarrow{n\to\infty}0\,,$$

同样它蕴含了 以接近于1的概率有$\frac{x_0x_k+\cdots+x_nx_{n+k}}{n+1}\approx R(k)$. 记

$$\hat{m} = \frac{x_0 + \cdots + x_n}{n+1}, \quad \hat{R}(k) = \frac{x_0 x_k + \cdots + x_n x_{n+k}}{n+1}. \qquad (11.6)$$

它们分别是均值 m 与相关函数 $B(k)$ 的最简便的相合估计.

显见, Gauss 宽平稳序列一定是平稳的. 对于具有谱密度的宽平稳 Gauss 列, 可以有更强的结论:

$$P(\frac{x_0 + \cdots + x_n}{n+1} \to m) = 1, \quad P(\frac{x_0 x_k + \cdots + x_n x_{n+k}}{n+1} \to R(k)) = 1.$$

相当一般的具有谱密度的宽平稳序 列 (例如, 只要满足 $\int_{-p}^{p} \ln f(l) dl > -\infty$) 都可以用下面 5.3 段中的 ARMA 模型来近似, 这就是在实际领域中 ARMA 模型被广泛采用的原因.

4.2 渐近平稳序列与渐近宽平稳序列

定义11.18 随机变量列 $x_n (-\infty < n < \infty)$ 称为渐近平稳序列, 如果对于任意 m, k,当 $t \to \infty$ 时 $(x_{t+1}, \cdots, x_{t+m})$ 与 $(x_{t+k+1}, \cdots, x_{t+k+m}\}$ 的分布都依分布收敛到相同的分布. 而渐近平稳序列正是指在时间充分发展后近似平稳的时间序列. 随机变量列 $x_n (-\infty < n < \infty)$ 称为渐近宽平稳序列, 如果对于任意 n, k, 存在常数 m 及函数 $R(k)$, 使

$$\lim_{t \to \infty} Ex_{t+n} = m, \lim_{t \to \infty} E(x_{t+n} x_{t+n+k}) = R(k),$$

m 称为渐近均值, $R(k)$ 称为渐近相关函数. 而渐近宽平稳序列正是指在时间充分发展后近似宽平稳的时间序列.

例11.19 具有不变分布的不可约 Markov 链, 在初始值服从不变分布时, 是平稳序列; 而当初始值任意时, 是渐近平稳序列.

在实际数据处理应用中, 用宽平稳序列来拟合建模的常常只是渐近宽平稳序列. 然而这在应用中已经足够了, 因为 我们都可以认为当前的时刻已经是该时间序列已发展到达到了充分长的阶段. 也就是说, 在实用中我们并不严格区分宽平稳序列与渐近宽平稳序列

4.3 平稳增量序列

如果随机序列显著地出现一个随时间发展的趋势, 则在模型拟合时, 可以研究其增量: $h_n \overset{\Delta}{=} x_n - x_{n-1}$, 对于此增量序列, 考察其平稳性, 宽平稳性, 渐近平稳性与渐近宽平稳性, 并进一步选择合适的模型, 以便得到较为合适的模型参数拟合. 宽平稳增量序列经济学家们常称为单位根过程

[注] $E(x_n x_{n+k}) = R(k), Ex_n = 0$ 的平稳列 $x_n (n = 1, \cdots, N)$ 的样本可作如下的模拟.

设 $R(k)$ 满足: 对于任意 l, 任意 $n_1, \cdots, n_l, (R(n_i - n_j))_{ij \le l}$ 都是一个非负定的矩阵. 取独立同分布列 $\{w_n\}$ 使 $Ew_n = 0, Ew_n^2 = R(0)$, 待定线性组合 $x_k = c_1 w_k. + \cdots + c_n w_{k+n-1}$, 其系数 $(c_1, \cdots, c_n)$ 可由方程组

$$R(k-1) = E[(\sum_{i=1}^{N} c_i w_{k+i-1})(\sum_{j=1}^{N} c_j w_j)] = R(0)[c_1 c_k + c_2 c_{k+1} + \cdots + c_{n-k+1} c_n)$$

解得. 这是一个二次方程组, 可以用模拟退火方法.

5. ARMA模型 (Auto-Regression Moving Average 模型)

5.1 ARMA (p, q)

定义11.20 对于宽平稳序列 x_n, 若存在不相关的时间序列 $e_n (-\infty < n < \infty)$ (在实用中还常常假定它们是独立同分布的, 这时 x_n 还是平稳序列), 使 $Ee_n = 0, Vare_n = s^2$, 且

$$x_n - a_1 x_{n-1} - \cdots - a_p x_{n-p} = b_0 e_n + b_1 e_{n-1} + \cdots + b_q e_{n-q}, \qquad (11.7)$$

其中多项式 $A(z) = 1 - a_1 z - \cdots - a_{p-1} z^{p-1} - a_p z^p$ 在 $|z| \le 1$ 无零点，(这个条件保证了 x_n 的稳定性质, 即渐近宽平稳性, 也就是经过连续地递推后, 不会趋于无穷)，则称 $x_n (-\infty < n < \infty)$ 为 (p，q) 阶 ARMA 模型. 记为 ARMA(p,q).

在实用中, 由于宽平稳性, 上面的系数 $(a_1, \cdots, a_p), (b_0, b_1, \cdots, b_q)$ 可以通过序列 x_n 的一段观测值来估计.

不难证明，宽平稳序列为 ARMA (p, q)的充要条件是它具有如下形式的谱密度

$$f(l) = C\frac{|B(e^{il})|^2}{|A(e^{il})|^2}, \qquad (11.8)$$

其中多项式 $A(z) = 1 - a_1 z - \cdots - a_{p-1} z^{p-1} - a_p z^p$ 在 $|z| \le 1$ 无零点，而 $B(z) = b_0 + b_1 z + \cdots + b_{q-1} z^{q-1} + b_q z^q$.

满足 $(b_1, \cdots, b_q) = 0$ 的 ARMA模型，简记为 AR(p)，称为 p 阶自回归模型，或 p 阶宽马氏模型；而满足 $(a_1, \cdots, a_p) = 0$ 的 ARMA模型 ，简记为 MA(q), 称为 q 阶滑动平均模型.

在应用学科中常遇到渐近宽平稳序列，实际上可以把它当作宽平稳序列, 并用 ARMA 模型作数据拟合．就是把数据列看成某个 ARMA 序列的一段样本．对于看起来具有平稳性质的数据列，一般地想象，似乎被拟合的 ARMA 模型的阶 p, q 越大越自由．事实却不是这样．过大而不适当的 p,q 不仅会增加了大量冗余计算工作量，有时还会出现过分拟合(over-fitting), 这同样会带来误差．所以，寻找尽量小的 p, q , 并使之能得到可以接受的近似, 称为模型的定阶问题．

实用中的 AR(p), 常常假定其自回归残差 e_n 是独立的随机变量序列, 所以在 $x_{n-1}, \cdots, x_{n-p}$ 已知的条件下, x_n 的条件期望为

$$\begin{aligned} E(x_n | (x_{n-1}, \cdots, x_{n-p})) &= E(([a_1 x_{n-1} + \cdots + a_p x_{n-p}] + b_0 e_n) | (x_{n-1}, \cdots, x_{n-p})) \\ &= a_1 x_{n-1} + \cdots + a_p x_{n-p} \end{aligned}. \qquad (11.8)$$

可见，知道了自回归模型系数，就可以用时刻 n 以前的资料，去预测时刻 n 时的估值，或以多大的概率在什么范围内变化（区间估计）．如果还假定 e_n 服从正态分布，那么上式说明，对于模型 AR(p),在 $x_{n-1}, \cdots, x_{n-p}$ 已知的条件下, x_n 的条件分布为正态分布:

$$N(a_1 x_{n-1} + \cdots + a_p x_{n-p}, b_0^2 s^2),$$

(可以把 $(b_0 s)^2$ 合为一个参数来估计，这等价于令 $b_0 = 1$，所以以后对于 AR 模型，我们不妨假定 $b_0 = 1$)，这时还可以得到误差的区间估计．

5．2　AR 模型的定阶与偏相关系数以及模型参数的估计

定义１１．２１　设 x_n 为宽平稳序列 (并不限于 ARMA模型)．若实数 $a_j^{(k)} (j \le k)$ 满足

$$E|x_n - \sum_{j=1}^{k} a_j^{(k)} x_{n-j}|^2 = \inf_{c_1, \cdots, c_k} E|x_n - \sum_{j=1}^{k} c_j x_{n-j}|^2, \qquad (11.9)$$

则 $a_k^{(k)}$ 称为第 k 个偏相关系数．

用实测数据来拟合 AR 模型时，首先要判断用 AR 模型近似是否合适; 其次要估计 p (定阶); 最后还要估计 AR(p)模型的 p+1 个待估参数．以下的准则可以由简单的计算直接验证：

准则１１．２２　x_n 为 AR(p)，当且仅当，第 p 个偏相关系数非零, 而以后的偏相关系数都是零．

偏相关系数的求法 记 $F_k(c_1,\cdots,c_k)=E|x_n-\sum_{j=1}^{k}c_jx_{n-j}|^2$. 由定义它在 $(a_1^{(k)},\cdots,a_k^{(k)})$ 处取最小, 所以，粗略地，$(a_1^{(k)},\cdots,a_k^{(k)})$ 是方程组: $\frac{\partial F_k}{\partial c_j}=0(j=1,\cdots,k)$ 的解（它们还有可能是局部极值点而非整体极小）. 这个方程组也就是:

$$\begin{pmatrix} R(0) & R(1) & \cdots & R(k-1) \\ R(1) & R(0) & \cdots & R(k-2) \\ \vdots & \vdots & \vdots & \vdots \\ R(k-1) & R(k-2) & \cdots & R(0) \end{pmatrix}\begin{pmatrix} a_1^{(k)} \\ a_2^{(k)} \\ \vdots \\ a_k^{(k)} \end{pmatrix}=\begin{pmatrix} R(1) \\ R(2) \\ \vdots \\ R(k) \end{pmatrix}, \qquad (11.10)$$

称为 Yule-Walker 方程. 由此可以解出第 k 个偏相关系数 $a_k^{(k)}$ 的理论值. 在实际情形, 我们只知道一段样本 $x_1,\cdots,x_N$, 由此可以先得到 $R(j)(j\le m)$ 的估计 $\hat{R}(j)(j\le m)$. 再用它们来代替上面计算第 k 个偏相关系数 $a_k^{(k)}$ 的 Yule-Walker 方程中的 $R(j)(j\le k)$, 得到的方程称为经验 Yule-Walker 方程, 它的解便是第 k 个偏相关系数 $a_k^{(k)}$ 的估计 $\hat{a}_k^{(k)}$, 称为经验偏相关系数.

检查多项式 $1-a_1z-\cdots-a_{p-1}z^{p-1}-a_pz^p$ 在 $|z|\le 1$ 有无零点, 在实践中是非常困难的. 通常的方法在实用中既复杂且很难实际操作. 因此，人们在拟合 AR(p)模型时常常加上约束条件 $\sum_{i=1}^{p}|a_i|<1$，以保证多项式 $1-a_1z-\cdots-a_{p-1}z^{p-1}-a_pz^p$ 在 $|z|\le 1$ 无零点. 这样就使估计参数问题实际成为求解一个在约束条件下的最小值问题.

偏相关系数的递推算法

由(11.10) 利用归纳法, 可以得到偏相关系数的递推算法：为此我们归纳地假定在 k 时已经有了表达式. 而对 $k+1$ 情形，可以由 $k+1$ 阶 Yule-Walker 方程定义如下递推方程

$$\begin{pmatrix} R(0) & R(1) & \cdots & R(k) \\ \vdots & \vdots & \cdots & \vdots \\ R(k-1) & R(k-2) & \cdots & R(1) \\ R(k) & R(k-1) & \cdots & R(0) \end{pmatrix}\begin{pmatrix} a_1^{(k+1)} \\ \vdots \\ a_k^{(k+1)} \\ a_{k+1}^{(k+1)} \end{pmatrix}=\begin{pmatrix} R(1) \\ \vdots \\ R(k) \\ R(k+1) \end{pmatrix}.$$

递推地求解这个方程，就可得到 $\{a_j^{(k+1)},(j\le k+1)\}$ 与 $\{a_j^{(k)},(j\le k\}$ 间的递推关系.

下面推导此方程的递推解法. 注意此方程的系数是一个 Toepliz 矩阵, 以 Toeolitz 矩阵为系数的线性方程组的解法是非常经典的. 首先, 我们把它改写为矩阵方程. 记系数矩阵为 $\mathbf{R_k}$, 它是对称的而且对于 $ARMA$ 模型易证它是正定的 (作为习题). 再记 $\mathbf{T_k}$ 为如下的 k 阶倒向算子：

$$\mathbf{T_k}=\begin{pmatrix} & & 1 \\ & \iddots & \\ 1 & & \end{pmatrix}, \mathbf{r_k}=\begin{pmatrix} R(1) \\ \vdots \\ R(k) \end{pmatrix}, \mathbf{x_k}=\begin{pmatrix} a_1^{(k)} \\ \vdots \\ a_{k-1}^{(k)} \end{pmatrix}.$$

于是对应的矩阵方程可写为

$$\begin{bmatrix} \mathbf{R_k} & \mathbf{T_k r_k} \\ \mathbf{r_k^T T_k^T} & \mathbf{R}(0) \end{bmatrix}\begin{bmatrix} \mathbf{x_{k+1}} \\ a_{\mathbf{k+1}}^{(\mathbf{k+1})} \end{bmatrix}=\begin{bmatrix} \mathbf{r_k} \\ \mathbf{R(k+1)} \end{bmatrix}.$$

也就是

$$\mathbf{R_k x_{k+1}} + \mathbf{T_k r_k} a_{\mathbf{k+1}}^{(\mathbf{k+1})} = \mathbf{r_k},$$

$$\mathbf{r_k^T T_k^T x_{k+1}} + \mathbf{R}(0) a_{\mathbf{k+1}}^{(\mathbf{k+1})} = \mathbf{R}(\mathbf{k}+1).$$

注意$\mathbf{R_k}$与$\mathbf{T_k}$是相互交换的, 由前一个方程, 利用$\mathbf{R_k^{-1} r_k} = \begin{bmatrix} \mathbf{x_k} \\ a_{\mathbf{k}}^{(\mathbf{k})} \end{bmatrix}$, 便有

$$\mathbf{x_{k+1}} = \mathbf{R_k^{-1} r_k} - a_{\mathbf{k+1}}^{(\mathbf{k+1})} \mathbf{T_k R_k^{-1} r_k} = \begin{bmatrix} \mathbf{x_k} \\ a_{\mathbf{k}}^{(\mathbf{k})} \end{bmatrix} - a_{\mathbf{k+1}}^{(\mathbf{k+1})} \begin{bmatrix} a_{\mathbf{k}}^{(\mathbf{k})} \\ \mathbf{T_{k-i} x_k} \end{bmatrix}, \qquad (11.11\text{-}1)$$

把它代入后一个方程得到

$$\mathbf{R}(\mathbf{k}+1) = \mathbf{r_k^T T_k^T} (\mathbf{R_k^{-1} r_k} - a_{\mathbf{k+1}}^{(\mathbf{k+1})} \mathbf{T_k R_k^{-1} r_k}) + a_{\mathbf{k+1}}^{(\mathbf{k+1})} \mathbf{R}(0)$$

$$= a_{\mathbf{k+1}}^{(\mathbf{k+1})} (\mathbf{R}(0) - \mathbf{r_k^T R_k^{-1} r_k}) + \mathbf{r_k^T T_k^T R_k^{-1} r_k}.$$

由此推出

$$a_{\mathbf{k+1}}^{(\mathbf{k+1})} = \frac{\mathbf{R}(\mathbf{k}+1) - \mathbf{r_k^T T_k^T} \begin{bmatrix} \mathbf{x_k} \\ a_{\mathbf{k}}^{(\mathbf{k})} \end{bmatrix}}{\mathbf{R}(0) - \mathbf{r_k^T} \begin{bmatrix} \mathbf{x_k} \\ a_{\mathbf{k}}^{(\mathbf{k})} \end{bmatrix}} = \frac{R(k+1) - \sum_{j=1}^{k} R(k+1-j) a_j^{(k)}}{R(0) - \sum_{j=1}^{k} R(j) a_j^{(k)}}. \qquad (11.11\text{-}2)$$

综合(11.11-1)与(11.11-2), 就是理论偏相关系数列的以下的递推算法：

$$\begin{cases} a_1^{(1)} = \dfrac{R(1)}{R(0)} \\ a_2^{(2)} = \dfrac{R(2) - R(1) a_1^{(1)}}{R(0) - R(1) a_1^{(1)}} \\ \quad \cdots \\ a_{k+1}^{(k+1)} = \dfrac{R(k+1) - \sum_{j=1}^{k} R(k+1-j) a_j^{(k)}}{R(0) - \sum_{j=1}^{k} R(j) a_j^{(k)}} \\ a_j^{(k+1)} = a_j^{(k)} - a_{k+1}^{(k+1)} a_{k-(j-1)}^{(k)} \qquad (j \le k) \end{cases} \qquad (11.11)$$

同样, 如果用$\hat{R}(j)(j \le m)$代替$R(j)(j \le m)$, 则就得到相应的估计$\hat{a}_j^{(k)}(j \le k)$.

由数据对 AR 模型粗定阶

对于一段样本$x_1, \cdots, x_N$, 要拟合 AR 模型, 则首先要确定 AR 模型的阶. 最简单的想法是: 逐个地计算出偏相关系数, 如果对于某个k而言, 以后的$a_n^{(n)}(n > k)$已经达到实际地足够小, 则可以近似地认为$p = k$.　但是这个方法既粗糙且在实际中显然并非可行．而更为实用的是后面将要介绍的 AIC 定阶法与 BIC 定阶法（参见 5．4 段中的注 1）．

AR(p)的相关序列的 Yule-Walker 方程与自回归系数$(a_1, \cdots, a_p)$的估计

对于 AR(p) 模型的自回归系数, 自然地采用如下估计

$$\hat{a}_i = \hat{a}_i^{(p)} (i \le p)$$

另一种看法是：注意 AR(p)的相关序列$\{R(k)\}$满足以下的 Yule-Walker 方程:

$$R(k)=a_1R(k-1)+\cdots+a_pR(0),(k\geq 1)\,. \qquad (11.\ 10)'$$

这个方程与(11. 10)**是一样的, 其差别仅在于**: (11. 10)**是为了用求来偏相关系数及定阶, 所以未知数实际上有无穷多个, 而如今假定已知其阶为** p **, 未知数就只有** p **个, 方程虽然有无穷多个, 却只有** p **个是线性无关的．所以是有限个方程．又显见**(11. 10)' **的解恰好是此时间序列的前** p **个偏相关系数．只要在(11．10)' 中用** $\hat{R}(j)(j\leq p)$ **代替** $R(j)(j\leq p)$ **，就得到相应的自回归系数** $(a_1,\cdots,a_p)$ **的如下估计** $\hat{\mathbf{a}}\overset{\Delta}{=}(\hat{a}_1,\cdots,\hat{a}_p)^T$ **：**

$$\hat{\mathbf{a}}=\hat{\mathbf{R}}^{-1}\hat{\mathbf{b}}\,,$$

其中

$$\hat{\mathbf{R}}=\begin{pmatrix}\hat{g}_0 & \hat{g}_1 & \cdots & \hat{g}_{p-1}\\ \hat{g}_1 & \hat{g}_0 & \cdots & \hat{g}_{p-2}\\ \cdots & \cdots & \cdots & \cdots\\ \hat{g}_{p-1} & \hat{g}_{p-1} & \cdots & \hat{g}_0\end{pmatrix},\ \hat{\mathbf{b}}=(\hat{g}_1,\hat{g}_2,\cdots\hat{g}_p)^T\,.$$

这个估计一般称为 Yule-Walker 估计．实际上，它正好是前面得到的估计 $\hat{a}_i=\hat{a}_i^{(p)}(i\leq p)$.

此外, 也可以用最小二乘估计, 即令

$$\dot{x}=\begin{pmatrix}x_N\\ x_{N-1}\\ \vdots\\ x_{p+1}\end{pmatrix},A=\begin{pmatrix}x_{N-1} & \cdots & x_{N-p}\\ x_{N-2} & \cdots & x_{N-p-2}\\ \vdots & & \vdots\\ x_p & \cdots & x_1\end{pmatrix},N>>p,\dot{q}=\begin{pmatrix}a_1\\ a_2\\ \vdots\\ a_p\end{pmatrix}. \qquad (11.12)$$

用方程 $A\vec{q}=\vec{x}$ **在** $|a_1|+\cdots+|a_p|<1$ **的条件下的约束最小二乘解 (参见第一章), 作为自回归系数** $\vec{q}$ **的估计. (最粗略的处理约束方法是对建模误差加上一个惩罚项, 例如** $C(\sum_{k=1}^{p}|a_k|-1)^+$ **, 其中** C **是一个很大的正数, 上标的符号表示取正部, 即** $a^+=aI_{[0,\infty)}(a)$).

残差方差的估计

由 $x_n-a_1x_{n-1}-\cdots-a_px_{n-p}=e_n$ **, 立刻可以残差的方差** $s^2=Ee_n^2$ **的估计**

$$\hat{s}^2=\frac{1}{N-p}\sum_{n=p+1}^{N}(x_n-\hat{a}_1x_{n-1}-\cdots-\hat{a}_px_{n-p})^2=\hat{R}(0)-\hat{\mathbf{a}}^T\hat{\mathbf{b}}=\hat{\mathbf{R}}(0)-\hat{\mathbf{b}}\hat{\mathbf{R}}^{-1}\hat{\mathbf{b}}\,.$$

［注 1］ 如前所述, 如果自回归残差 e_n **服从正态分布, 就可以对自回归系数及条件方差作最大似然估计, 进而还可以对它们作区间估计. 在应用中一般地自回归残差并不服从正态分布, 但是如果无视是否有正态性, 强行按正态分布作最大似然估计与区间估计, 也能得到较粗的参考信息.**

[注 2] **与统计领域中不同，在经济类的一些有关时间序列的书籍与文献中，**11. 7) **代之以**

$$x_n-a_0-a_1x_{n-1}-\cdots-a_px_{n-p}=b_0e_n+b_1e_{n-1}+\cdots+b_qe_{n-q}\,, \qquad (11.\ 7)^*$$

并且也称之为 ARMA(p,q)**模型。这种扩展增加了一个常数项** a_0 **，这种模型出自经济研究．当** $a_0>0,a_1\geq 0,\cdots,a_p\geq 0,a_1+\cdots+a_p<1$ **时, 可以证明此时间序列是渐近平稳的．**

[**注** 3] **由理论上可以证明, 在给定相关函数列的** $p+1$ **个值** $R(0),R(1),\cdots,R(p)$ **的所有平稳序列中, 使**

$$c_0^2 \overset{\Delta}{=} \inf_{k, g_1, \cdots, g_k} E|x_n - \sum_{j=1}^{k} g_j x_{n-j}|^2$$

达到极大的平稳序列是 AR(p)模型. 这里的 c_0 称为平稳序列的一步预测的(均方)误差. 这就是说, AR(p)模型给出了在这种限制下, 平稳序列的一步预测的误差的上界，即给出了最坏的可能. 此外，人们还证明了几乎在同样的条件下, 在具有谱密度的所有平稳序列中, 只有正态 AR(p)平稳序列 $\{x_n\}$ 满足： 对于任意 $N > p$，$(x_1, \cdots x_N)$ 的联合分布的熵最大. 这也给出了不确定性最大时的建模, 也最坏的可能.

5．3　MA 模型的定阶与参数估计

容易验证以下准则：

准则１１．２３　宽平稳序列 x_n 为 MA(q), 当且仅当, $R(q) \neq 0, R(k) = 0(|k| \geq q)$.

由此可以通过由实测数据估计相关系数 $R(k)$， 观察它的大小， 以给出 q 的粗估计. 但是更为使用的是在下面 5．4 段注 1 中的 AIC 估计法与 BIC 估计法.

滑动平均系数 $(b_1, \cdots, b_q)$ 的粗估计

在假定 q 已估计得到的前提下, 由模型的定义有 $R(k) = s^2 \sum_{j=k}^{q} b_j b_{j-k}, (0 \leq k \leq q)$．由求解二次方程 $\hat{R}(k) = s^2 \sum_{j=k}^{q} x_j x_{j-k}, (0 \leq k \leq q)$ 得到的 $(x_0, x_1, \cdots, x_q)$，就是 $(b_0, b_1, \cdots, b_q)$ 的估计.

求解二次方程 $\hat{R}(k) = \sum_{j=k}^{q} b_j b_{j-k}, (0 \leq k \leq q)$，即求解非线性方程

$$\hat{R}(0) = \hat{s}^2 (b_0^2 + \cdots + b_q^2),$$

$$\hat{R}(k) = \hat{s}^2 (b_0 b_k + b_1 b_{k+1} \cdots + b_{q-k} b_q), (1 \leq k < q), \quad \hat{R}(q) = \hat{s}^2 b_0 b_q, \quad (11.12)$$

此方程组的解可由如下的迭代近似:

$$\hat{R}(0) = \hat{s}^2 [(b_0^{(n+1)})^2 + (b_1^{(n)})^2 \cdots + (b_q^{(n)})^2],$$

$$\hat{R}(k) = \hat{s}^2 [b_0^{(n+1)} b_k^{(n+1)} + b_1^{(n)} b_{k+1}^{(n)} \cdots + b_{q-k}^{(n)} b_q^{n)}], (1 \leq k < q), \hat{R}(q) = \hat{s}^2 b_0^{(n+1)} b_q^{(n+1)}$$

得到．另外，他们也可由求解优化问题 $\inf_{b_0, b_1, \cdots, b_q} \sum_{k=1}^{q} |\hat{R}(k) - s^2 \sum_{j=k}^{q} b_j b_{j-k}|^2$ 得到近似值．不幸的是这些方法都过于繁琐， 在实际计算时，只在 $q < 5$ 时作数值求解才比较合适．

MA 模型估计的经验实际操作常常是，先拟合一个 AR 模型，得到残差的粗估计，然后把 MA 模型的估计简化为求解一个线性方程组．这是一种纯粹经验性的方法，只有直观根据．其具体操作如下：

假定由一段样本 $x_1, \cdots, x_N$，经过下面的 5．4 段注 1 中的 AIC 或 BIC 法已经估计好了 MA 模型的阶 q．取 s 满足 $q << s << N$．先用此数据拟合一个 AR(s)模型. 由这个拟合了的模型, 算出残差的估计 $\hat{e}_{s+1}, \cdots, \hat{e}_N$ (其中 $\hat{e}_k = x_k - (\hat{a}_0 x_{k-1} + \cdots + \hat{a}_s x_{k-s})$). 最后求方程

$$x_k = -(b_0 \hat{e}_k + \cdots + b_q \hat{e}_{k-q})$$

的最小二乘解 $(b_0, b_1, \cdots, b_q)$，作为 MA(q)的模型参数估计．

5．4　ARMA**模型的定阶与参数估计**

对于 ARMA(p,q)

$$x_n - a_1 x_{n-1} - \cdots - a_p x_{n-p} = b_0 e_n + b_1 e_{n-1} + \cdots + b_q e_{n-q}$$

的粗定阶与估计，**我们有**

准则１１．２４ 宽平稳序列x_n**为** ARMA(p,q)，**当且仅当，存在多项式** $A(z) = 1 - a_1 z - \cdots - a_{p-1} z^{p-1} - a_p z^p$ **在**$|z| \le 1$**无零点**，**且对于任意**$m > q$**有**

$$R(m) = \sum_{k=1}^{p} a_k R(m-k)\,. \qquad \text{（１１．１３）}$$

在条件满足时$(a_1, \cdots, a_p)$ **就是这个自回归滑动和的** AR **部分的系数．**

自回归系数$(a_1, \cdots, a_p)$**与滑动平均系数**$(b_0, b_1, \cdots, b_q)$ **的粗估计**

第 1 **步： 用宽平稳序列**x_n **的样本计算相关函数序列的估计**$\hat{R}(n)$**．**

第 2 步：对于$q = 1, 2, \cdots$，**看是否存在** $p \ge 1$**及充分大的**m **使** $\begin{pmatrix} \hat{R}(q+1) \\ \vdots \\ \hat{R}(q+m) \end{pmatrix}$**能近似地能由** $\begin{pmatrix} \hat{R}(q+1-k) \\ \vdots \\ \hat{R}(q+m-k) \end{pmatrix}$$(k \le p)$ **线性表出**，**如果对于某个** q(**取尽量小**)**存在这样的近似线性表出**， **则这样的**(p, q)**就可取为此** ARMA**模型的阶**，**而这个线性表出的系数就可以作为** $(a_1, \cdots, a_p)$**的估计** $(\hat{a}_1, \cdots, \hat{a}_p)$**．**

第 3 **步： 对于上面确定的**(p, q)**及**$(\hat{a}_1, \cdots, \hat{a}_p)$，**把下述 最小值问题**

$$\inf_{x_0, x_1, \cdots, x_q} \sum_{l=1}^{q} \left| \hat{R}(l) - \sum_{k=1}^{p} \hat{a}_k \hat{R}(l-k) - \sum_{j=m}^{q} x_j x_{j-m} \right|^2$$

取到最小值的位置$(x_1, \cdots, x_q)$**作为**$(b_0, b_1, \cdots, b_q)$**的估计**$(\hat{b}_0, \cdots, \hat{b}_q)$**．**

[**注** 1] **以上的第２步与第３步在实际中很难使用． 在实用中更常用的**ARMA **模型的经验定阶法有：**

（1） AIC（Akaike Information Criterion）**法** (**这是** Akaike**采用相对熵导出的定阶原则**)

设$x_1, \cdots, x_N$**是来自正态** ARMA**模型的一段观测值．对于指定的**k **（例如取可能的阶的一个上界，或者取它为与**$\ln N$ **同阶的一个整数）及固定的**$(p, q), (0 \le p, q \le k)$，**记相应的**$s^2$ **的最大似然估计为** $\hat{s}^2(p, q)$**．记**

$$AIC(p, q) = \ln \hat{s}^2(p, q) + \frac{2(p+q)}{N}\,.$$

取$(\hat{p}, \hat{q})$，**使**

$$AIC(\hat{p}, \hat{q}) = \min_{0 \le p, q \le k} AIC(p, q)\,。$$

那么可以用$(\hat{p},\hat{q})$作为阶的估计．这就是 AIC方法． 虽然这个定阶法在 ARMA 模型的定阶中常用, 但是却被证明了它并不是阶的相容估计, 即人们已经证明了，如果样本来自 $ARMA(p_0,q_0)$, 那么就有

$$P(\hat{p}< p_0,\hat{q}<q_0)=0，P(\hat{p}> p_0\text{或}\hat{q}>q_0)>0。$$

从而说明用 AIC方法得到的阶可能比 ARMA 模型的真正的阶要大.

（2） BIC, HIC 法 (AIC的改进)

用下面的 BIC (Bayesian Information Criterion), HIC 代替 AIC

$$BIC(p,q)=\ln \hat{s}^2(p,q)+\frac{(p+q)\ln N}{N},$$

$$(0\le p,q\le k),,HIC(p,q)=\ln \hat{s}^2(p,q)+\frac{c(p+q)\ln\ln N}{N},(c>2).$$

如果我们把对应于样本$x_1,\cdots,x_N$ 的 BIC估计记为 $(\hat{p}_N,\hat{q}_N)$, 那么在$x_1,\cdots,x_N$ 为 $ARMA(p_0,q_0)$ 时, 已经证明了当$N\to\infty$ 时，依概率地有 $\hat{p}_N\to p_0,\hat{q}_N\to q_0$． 这说明 BIC 估计是相合的．

在作线性预报时, 一般地并不在乎高估其阶数, 所以常用 AIC 估计． 但是，在对模型结构作研究时, 就应该用 BIC 估计．

ARMA 模型参数的估计

对于 ARMA(p,q), 易见有

$$R(q+k)=a_1R(q+k-1)+\cdots+a_pR(q+k-p),(k\ge1)\text{．}$$

由此可以从观测数据$(x_1,\cdots,x_N)$, 估计$R(q+k)$, 再解出$(a_1,\cdots,a_p)$, 得到估计.$(\hat{a}_1,\cdots,\hat{a}_p)$.
然后记

$$h_n=x_n-(a_1x_{n-1}+\cdots+a_px_{n-p})\text{．}$$

那么不难验证$\{h_n\}$是 MA(q)模型．最后 用h_n的估计 $\hat{h}_n=x_n-(\hat{a}_1x_{n-1}+\cdots+\hat{a}_px_{n-p})$代替$h_n$, 作为已知数据来拟合 MA(q), 得到模型参数$(b_0,b_1,\cdots,b_q)$的估计$(\hat{b}_0,\cdots,\hat{b}_q)$．

[注２] 由宽平稳序列的一般理论, 一个宽平稳序列可以分成两个部分, 一部分是其所有时刻的共有信息, 称为奇异部分，而另一部分则表达了随时刻变化所带来的新的信息, 称为正则部分． 在实际应用中, 人们通常认为宽平稳序列只有正则部分，此时不仅具有谱密度 $f(l)$, 而且还满足 $\int_0^{2p}\ln f(l)dl>-\infty$，这时自然就可以用 ARMA 模型近似. 这是 ARMA 模型被广泛应用的原因．

5．5 ARMA模型的预报问题

我们在这一段除假定残差序列$\{e_n\}$独立同分布, $A(z)=1-a_1z-\cdots-a_{p-1}z^{p-1}-a_pz^p$ 在$|z|\le1$ 无零点外, 还增加一个假定，姑且称之为假定 B：

$$B(z)=b_0+b_1z+\cdots+b_{q-1}z^{q-1}+b_qz^q\text{ 在}|z|\le1\text{ 也无零点．}$$

此时$\frac{A(z)}{B(z)},\frac{B(z)}{A(z)}$ 都是(复)解析函数, 我们把它门分别记为

$$\frac{B(z)}{A(z)}=\sum_{k=0}^{\infty}c_kz^k,\ \frac{A(z)}{B(z)}=\sum_{k=0}^{\infty}d_kz^k\text{．}$$

记D为如下的向后推移运算：

$$Dx_n=x_{n-1},De_n=e_{n-1}.$$

于是 ARMA(p,q)可以写为

$$A(D)x_n = B(D)e_n .$$

从而有

$$x_n = \sum_{k=0}^{\infty} c_k e_{n-k}, e_n = \sum_{k=0}^{\infty} d_k x_{n-k} . \qquad (11.14)$$

随机变量序列$\{e_n\}$在数学文献中称为 Wold 序列, 而在控制领域中称为新息序列 ， 因为它对x_n带来了时间序列的过去 $\{x_{n-1},\cdots,x_1\}$所不能包含的新的信息. (11. 14)的第一个等式称为 Wold 分解，其中c_k称为第k个 Wold 系数.

在以上的假定下, 我们立刻得到下面的定理.

定理１１．２５ 对于满足假定 B 的正态 ARMA(p,q)模型(即$\{e_n\}$是独立同分布正态随机变量列), 存在常数$k_1,\cdots,k_n$ 使

$$E(x_{n+1} | x_n,\cdots x_1) = k_n x_n + \cdots + k_1 x_1 .$$

即最佳预报为线性函数.

作为正态 ARMA 模型的特例，我们有

１．正态 AR(p)的预报：

$$\hat{x}_{n+1} \overset{D}{=} E(x_{n+1} | x_n,\cdots x_1) = -(a_1 x_n + \cdots + a_p x_{n-p+1}), (n \ge p) .$$

$$\hat{x}_{n+2} = -(a_1 \hat{x}_{n+1} + a_2 x_n \cdots + a_p x_2), \cdots, \hat{x}_{n+p+1} = -(a_1 \hat{x}_{n+p} + \cdots + a_p \hat{x}_n)$$

２．正态 MA(q)的预报 ：

由(11. 14)可推出$E(e_{n+1} | x_n,\cdots x_1) = 0$, 因此

$$\hat{x}_{n+1} \overset{D}{=} E(x_{n+1} | x_n,\cdots x_1) = E[(b_1 e_n + \cdots + b_q e_{n-q+1}) | x_n,\cdots x_1]$$

$$= b_1 \sum_{k=0}^{\infty} d_k x_{n-k} + \cdots + b_q \sum_{k=0}^{\infty} d_k x_{n+1-q-k} \approx b_1 \sum_{k=0}^{n-1} d_k x_{n-k} + \cdots + b_q \sum_{k=0}^{n-q} d_k x_{n+1-q-k} .$$

$$\hat{x}_{n+k} = E(x_{n+k} | x_n,\cdots x_1) = 0, (k > q) .$$

类似地可得到 ARMA(p,q)的预报系数$k_1,\cdots,k_n$的表达式.

[注 1] 可以证明，对于 ARMA 模型，总可以经过改造使得$A(z), B(z)$ 在$|z|>1$上无零点． 所以，最为实质的假定是 $A(z), B(z)$ 在单位圆$|z|=1$上无零点． 当此条件不满足时, 模型所表达的时间序列就不再具有渐近平稳性． 这时，时间序列就会出现趋势项, 即是平稳增量过程.

[注 2] 在理论上可以证明, 一维 ARMA模型可以作为向量 AR 模型的一个分量． 因此 ARMA 模型的计算, 也可以归结为向量 AR 模型． 显然, 向量模型的计算也有其特殊的复杂之处，即矩阵乘法的不可交换性 .

[注 3] (AR模型的渐近平稳性的证明) 设$\{e_n\}$独立同分布, 且取值于有限区间,
$Ee_n = 0, Var(e_n) = s^2$

$$x_n - a_1 x_{n-1} - \cdots - a_p x_{n-p} = e_n ,$$

其中多项式$1 - a_1 z - \cdots - a_{p-1} z^{p-1} - a_p z^p$ 在$|z| \le 1$ 无零点. 令

$$h_n = \begin{pmatrix} x_n \\ x_{n-1} \\ \vdots \\ x_{n-p+1} \end{pmatrix}, V_n = \begin{pmatrix} e_n \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \; A = \begin{pmatrix} a_1 & \cdots & a_{p-1} & a_p \\ 1 & 0 & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{pmatrix}.$$

我们可把模型改写为

$$h_n = Ah_{n-1} + V_n . \tag{11. 15}$$

于是

$$h_n = Ah_{n-1} + V_n = A(Ah_{n-2} + V_{n-1}) + V_n$$
$$= \cdots = A^{n-p}h_{p-1} + V_n + AV_{n-1} + \cdots + A^{n-p}V_p) .$$

假定$(x_{p-1},\cdots,x_1,x_0)$与V_p独立, 利用$\{V_n\}$也是独立同分布的, 上式右方就与

$$A^{n-p}h_{p-1} + (A^{n-p}V_n + A^{n-p-1}V_{n-1} + \cdots + V_p) \overset{记}{=} I_n^{(1)} + I_n^{(2)}$$

有相同的分布. 由于A的特征多项式为$l^p - a_1 l^{p-1} - \cdots - a_{p-1} l - a_p$, 由 AR 模型的假定, 它 在$|\ l\ | \geq 1$无零点, 故$A$的 Jordan 分解可以写为

$$A = T\begin{pmatrix} \Lambda_1 & & \\ & \ddots & \\ & & \Lambda_l \end{pmatrix} T^{-1},$$ Λ_i为对角元为 l_i 的 Jordan 块, $|\ l_i\ | <$ 某$g < 1, (i \leq p)$.

于是有 $I_n^{(1)} \to 0$, $I_n^{(2)}$按概率$\to \sum_{m=p}^{\infty} A^{m-p}V_m$. 后一个极限的证明如下：首先, 按 L^2 的意义对 $N < M$ 有

$$\| \sum_{m=p}^{M} A^{m-p}V_m - \sum_{m=p}^{N} A^{m-p}V_m \|^2 = \sum_{m,n=N+1}^{M} tr(A^{m-p}E(V_m V_n^T)(A^T)^{n-p})$$
$$\leq \sum_{m,n=N+1}^{M} tr(A^{m-p}\frac{1}{2}[E(V_m V_m^T) + E(V_n V_n^T)](A^T)^{n-p})$$ (按非负定矩阵含义下的次序)
$$\leq \sum_{m,n=N+1}^{M} g^{m=n-2p}E(V_1 V_1^T) \leq [\sum_{m=N+1}^{M} g^{m-p}]^2 E(V_1 V_1^T) \to 0 .$$

即$\| I_n^{(2)} - \sum_{m=p}^{\infty} A^{m-p}V_m \| \to 0$. 再用 Chebyshev 不等式得到按概率的收敛性, 即

$$P(|\ I_n^{(2)} - \sum_{m=p}^{\infty} A^{m-p}V_m | > d) \leq \frac{1}{d^2} \| I_n^{(2)} - \sum_{m=p}^{\infty} A^{m-p}V_m \|^2 \to 0 .$$

从而h_n有不变分布函数, 所以它是渐近平稳的. 】

以上证明中实际上只用到了二阶矩计算, 所以关于$\{e_n\}$独立同分布的假定并不必要, 只需假定它们是同分布的不相关随机序列, 则结论仍然正确, 证明也只需作一点小小的改动.

6 ARCH 模型

6.1 ARCH(q)

ARCH 模型是随机方差列建模中最简单的一种模型. 它的含义为自回归条件异方差模型 (Auto-Regressive Conditional Heteoskedastic model).而 ARCH 正是这些英文单字第一个字母的拼写. 它是由 Engel 在 1982 年提出的一种时变条件方差模型, 其目的是想解除 AR 模型中自回归残差的方差为常数(即$Var(e_n) = s^2$)的限制，代之以只依赖残差列的过去$\{e_{n-1},\cdots,e_1\}$的一个随机序列.

$\mathrm{ARCH}(q)$ 描述的是某种随机误差序列误差 e_n （其典型例子是自回归模型的残差列）．在金融中的随机误差列常用条件方差的如下的 $\mathrm{ARCH}(q)$ (随机)模型描述: 在 $\{e_{n-1},\cdots,e_{n-q}\}$ 已知的条件下, e_n 的条件分布密度为

$$e_n \sim N(0, g_0 + g_1 e_{n-1}^2 + \cdots + g_q e_{n-q}^2), \qquad (g_0 > 0, g_1, \cdots, g_q \geq 0). \qquad (11.16)$$

（典型例子是"包容有发展趋势的自回归"模型残差：$e_n = x_n - (a_0 + a_1 x_{n-1} + \cdots a_p x_{n-p})$，其最简单情形为只有发展趋势项的"0 阶自回归模型" $e_n = x_n - a$，后者在经济序列中广为采用．这里 $a_0, a_1, \cdots, a_p$ 和 $g_0, g_1, \cdots, g_q$ 都是待估的模型参数. 当 $a_0 = g_1 = \cdots = g_q = 0$ 时, ARCH模型就退化为 AR 模型）.由于 ARCH模型引进了时变的随机条件方差, 这就使它比 AR 模型有更加广泛的包容性与适用性. 为了保证 $\{e_n^2\}$ 是渐近平稳随机序列, 我们简单地假定

$$g_1 + \cdots + g_q < 1 . \qquad (11.17)$$

$\mathrm{ARCH}(q)$ 模型也可表达为:

$$e_n = \sqrt{h_n} u_n, \ h_n = g_0 + g_1 e_{n-1}^2 + \cdots + g_q e_{n-q}^2, \qquad (11.16)'$$

其中 u_n 是独立同分布的 $N(0,1\}$ 随机变量列，且与 $(e_n, \cdots, e_{n-q})$ 独立.

$\mathrm{ARCH}(q)$ 模型也可以并不假定正态性, 此时可采取如下的模型

$$e_n^2 = g_0 + g_1 e_{n-1}^2 + \cdots + g_q e_{n-q}^2 + h_n, \qquad (11.18)$$

其中 $\{h_n\}$ 是独立同分布随机变量列, $Eh_n = 0, Var(h_n) = s^2$, 甚至还可以不要求 随机序列 $\{h_n\}$ 的相互独立性. 注意(11. 18)式的模型比(11.16)式更为一般．但是，用(11.18)式建模时缺乏条件分布的信息, 因此在估计模型参数时, 不能用最大似然方法. 另一方面, 只要注意到(11.18)式本身是一个 AR(q)模型, 所以参数估计可以套用 AR 模型的现成结果.

ARCH(q)的定阶与参数估计

对于 ARCH(q)模型的实际定阶，也可以用 AIC 估计与 BIC 估计．一旦阶确定以后，(11.16)式的参数估计可以设计如下, 令

$$p_i(e_i \mid e_{i-1}, \cdots, e_{i-q}; \vec{g}) = \frac{1}{\sqrt{2ph_i}} e^{\frac{e_i^2}{2h_i}}, \ L(\vec{g}) = \prod_{i=q+1}^{N} p_i(e_i \mid e_{i-1}, \cdots, e_{i-q}; \vec{g}), \qquad (11.19)$$

其中 $e_n = x_n - (a_0 + a_1 x_{n-1} + \cdots + a_p x_{n-q})$ 还含有参数 $\vec{a} = (a_0, a_1, \cdots, a_n)$．这里，$L(\vec{g})$ 是在 $\{e_q, \cdots, e_1\}$ 已知条件下的条件似然函数． 于是可以仿照最大似然估计方法，求 L 相应的最大值位置 $\hat{g}$，$\hat{a}$, 作为模型参数 $\vec{g}, \vec{a}$ 的估计．这里的计算归结为: 求一个条件最大值的位置. 这个位置可以用数值模拟方法来求得. 为了防止计算陷入局部极大，我们也可以使用模拟退火算法．

在实际拟合模型时， 如果发现从样本数据得到的随机误差 e_n 有较厚的尾部, 那么就不能用(11. 16) 式中的正态假定．此时除了用模型(11. 18)式外, 人们也常用 t 分布代替(11.16)式中的正态分布建模. 这时由于要求把方差作为参数写进分布密度, 所以我们需要事先把 $t(k)$ 分布的密度作一番修正. 我们在 $t(k)$ 分布的密度

$$\frac{\Gamma(\frac{k+1}{2})}{\sqrt{kp}\,\Gamma(\frac{k}{2})}(1 + \frac{x^2}{k})^{-\frac{k+1}{2}}$$

中增加引进一个影响方差的正值 ”重尾参数” a, 得到一个新的密度(本书中姑且称之为 $t(k,a)$ 分布):

$$f(x)=\frac{\Gamma(\frac{k+1}{2})}{\sqrt{kp}\,\Gamma(\frac{k}{2})}a^{-\frac{1}{2}}(1+\frac{x^2}{ak})^{-\frac{k+1}{2}}.$$

可以求得此 $t(k,a)$ 分布的方差为 $s^2 \overset{\Delta}{=} a\frac{k}{k-2}$. 在密度中将重尾参数用此分布的方差表示, 就得到

$$f(x)=\frac{\Gamma(\frac{k+1}{2})}{\sqrt{(k-2)p}\,\Gamma(\frac{k}{2})}\frac{1}{s}(1+\frac{kx^2}{s^2(k-2)})^{-\frac{k+1}{2}}.$$

仿照(11.16)式得到

$t(k,a)$ 分布下的 ARCH(q)模型: 在 $(e_{n-1},\cdots,e_{n-q})=(x_{n-1},\cdots,x_{n-q})$ 的条件下, e_n 的条件分布密度为

$$p_n(x_n\mid x_{n-1},\cdots,x_{n-q})=\frac{\Gamma(\frac{k+1}{2})}{\sqrt{(k-2)p}\,\Gamma(\frac{k}{2})}\frac{1}{\sqrt{h_n}}(1+\frac{kx^2}{h_n(k-2)})^{-\frac{k+1}{2}},$$

其中

$$h_n=g_0+g_1x_{n-1}^2+\cdots+g_p x_{n-q}^2,.$$

于是条件似然函数为

$$L(\dot{g})=\prod_{n=q+1}^{N}p_n(e_n\mid e_{n-1},\cdots,e_{n-q};\dot{g}).$$

从而类似地可以作最大似然估计.

6．3 ARCH(q)模型的方差的预报

设某种随机误差序列 e_n 已用 ARCH(q) 模型建模．那么，在 $\{e_{n-1},\cdots,e_1\}$ 已知的条件下, e_n 的条件期望为 0, 从而条件方差为

$$\begin{aligned}Var(e_n\mid e_{n-1},\cdots,e_{n-q}) &\overset{\Delta}{=} E(e_n^2\mid e_{n-1},\cdots,e_{n-q})-[E(e_n\mid e_{n-1},\cdots,e_{n-q})]^2\\ &=E(e_n^2\mid e_{n-1},\cdots,e_{n-q})\text{．}\end{aligned}$$

上式右方是 e_n^2 在 $\{e_{n-1},\cdots,e_{n-q}\}$ 已知的条件下的最佳一步预报， 我们把它记为 $\hat{e}_n^2$ ．又由于 u_n 与 $\{e_{n-1},\cdots,e_{n-q}\}$ 独立，且 $Eu_n^2=1$，故我们有

$$\begin{aligned}\hat{e}_n^2 &= E(e_n^2\mid e_{n-1},\cdots,e_{n-q})=E(u_n^2h_n\mid e_{n-1},\cdots,e_{n-q})\\ &=Eu_n^2E(h_n\mid e_{n-1},\cdots,e_{n-q})=g_0+g_1e_{n-1}^2+\cdots+g_qe_{n-q}^2\text{。}\end{aligned}$$

于是 e_n^2 的最佳一步预报公式可以重写为

$$\hat{e}_{n+1}^2 \overset{\Delta}{=} E(e_{n+1}^2\mid e_n,\cdots,e_{n-q+1})=Var(e_{n+1}\mid e_n,\cdots,e_{n-q+1})=g_0+g_1e_n^2+\cdots+g_qe_{n-q+1}^2\text{．}$$

类似地，e_n^2 的最佳两步预报为

$$\hat{e}_{n+2}^2 = Var(e_{n+2}^2 \mid e_n, \cdots, e_{n-q+1}) = E(e_{n+2}^2 \mid e_n, \cdots, e_{n-q+1}) = E(u_{n+2}^2 h_{n+2} \mid e_n, \cdots, e_{n-q+1})$$
$$= Eu_{n+2}^2 E(h_{n+2} \mid e_{n-1}, \cdots, e_{n-q}) = E[(g_0 + g_1 e_{n+1}^2 + \cdots + g_q e_{n-q+2}^2) \mid e_n, \cdots, e_{n-q+1}]$$
$$= g_0 + g_1 \hat{e}_{n+1}^2 + e_n^2 + \cdots + g_q e_{n-q+2}^2 .$$

进一步可得到多步预报的递推公式

$$\hat{e}_{n+k}^2 = g_0 + g_1 \hat{e}_{n+k=1}^2 + \cdots + g_q \hat{e}_{n+k-q}^2 .$$

*7 GARCH (p,q) 模型与其它随机方差模型

7．1 GARCH 模型

GARCH 模型是 ARCH**模型的一种推广，它的含义为广义自回归条件异方差模型** (Generalized Auto-Regressive Conditional Heteoskedastic model). GARCH**是这些英文单字第一个字母拼写**. GARCH **模型常被用来描述金融证券的随机波动率**(**单位时间收益率的方差**).

我们采用正态假定下的随机误差序列$\{e_n\}$ **的** GARCH(p,q) **模型，即假定在**$e_{n-1}, \cdots, e_{n-q}$ **已知的条件下**, e_n **的条件分布密度为**$N(0, h_n)$, **其中随机序列** $\{h_n\}$ **满足**

$$h_n - (b_1 h_{n-1} + \cdots + b_p h_{n-p}) = g_0 + g_1 e_{n-1}^2 + \cdots + g_q e_{n-q}^2, \quad (11.20)$$

这里 $b_1, \cdots, b_p \geq 0$,$g_0 > 0, g_1, \cdots, g_q \geq 0$ **还有初始参数** $h_1, \cdots, h_p > 0$ **都是待估参数．而**. GARCH(0,q) **就退化为** ARCH(q)．**可以证明** GARCH(p,q) **模型在条件**

$$\sum_{j=1}^{p} b_j , \sum_{k=1}^{q} g_k < 1$$

下是渐近平稳的时间序列．

GARCH **模型也常写成为**

$$e_n = u_n \sqrt{h_n}, \quad (11.21)$$

其中h_n **满足**(11. 20)**式**, $u_n \sim N(0,1)$, **且**u_n **与**$(e_{n-1}, \cdots, e_{n-q})$ **独立．**

7．2 金融证券模型中的 GARCH(1,1)

以下我们以金融证券中较为常用的 GARCH**模型为例**, **来看应如何估计这些模型参数**. **在金融证券中**, **记标的变量**(underlying variable, **常见的标的变量有股票**, **债券**, **随机利率等等**)**在时刻** t **的值为** S_t. **这是依时间参数**t **的一族随机变量**, **也就是一个随机过程**. **在著名的** Black-Scholes **模型中有一个基本假设**, **就是假设** $\ln S_t$ **为独立增量过程**. **我们对于标的证券进行间隔为**Δt **的离散采样**, **并把** $S_{i\Delta t}$ **简单地记为**S_i . **差商** $\frac{1}{\Delta t}(\ln S_{i+1} - \ln S_i)$ **称为标的证券在**$i\Delta t$ **时刻的瞬时收益率**, **记成** r_i (**于是有** $S_{i+1} = S_i e^{r_i \Delta t}$, **这是仿照按连续复利率**$r$ **的情形下**, **在** $i\Delta t$ **时刻的本金** P_i **到**$(i+1)\Delta t$ **时刻实际增值为**$P_{i+1} = P_i e^{r\Delta t}$ **而定义的**). Black-Scholes **模型假定了** $r_i \sim N(\tilde{m}, s^2)$, **其中** s **是一个正常数**, **称为波动率**(volatility), **它对于由证券的风险所带来的风险收益起到了实质的影响**. **由于在实际市场中证券收益的实证值与** Black-Scholes **模型给出的值有很显著的差异**, **人们认识到应该用一个随时间变化的随机的波动率**$\sqrt{h_i}$, **去代替** Black-Scholes **模型中的常数波动率** s ．**于是想到对收益率**r_i **与其平均值** $\tilde{m}$ **的偏差**$e_i \stackrel{\Delta}{=} r_i - \tilde{m}$ **用** GARCH**建模．记在**$(r_0, \cdots, r_{i-1})$ **已知的条件下偏差**e_i **的条件方差为**h_i ．**假定采用** Garch(1,1)**模型，即假定方差** h_n **满足**

$$h_i - b \cdot h_{i-1} = g + a e_{i-1}^2,$$

其中$\tilde{m} > 0, 0 \le a, b, a + b, g < 1$以及初始参数$h_0$ 都看成待估的模型参数. 它们可以通过样本$\{S_i : i \le N\}$用最大似然估计法得到. 于是, 我们有条件异方差的如下表达式

$$\begin{aligned} h_{i+1} &= g + a(r_i - \tilde{m})^2 + bh_i = g + a(r_i - \tilde{m})^2 + b(g + a(r_{i-1} - \tilde{m})^2 + bh_{i-1}) \\ &= g(1+b) + a[(r_i - \tilde{m})^2 + b(r_{i-1} - \tilde{m})^2] + b^2 h_{i-1} = \cdots \\ &= g(1 + b + \cdots + b^i) + a[(r_i - \tilde{m})^2 + b(r_{i-1} - \tilde{m})^2 + \cdots + b^i (r_0 - \tilde{m})^2] + b^{i+1} h_0 \\ &\stackrel{\Delta}{=} h_{i+1}(r_i, \cdots, r_0; \tilde{m}, a, b, g, h_0\} . \end{aligned}$$

参数的估计可进行如下：对于$i \ge 1$, 记条件分布为

$$p_i(r_i \mid r_{i-1} \cdots, r_0; \tilde{m}, a, b, g, h_0) = \frac{1}{\sqrt{2p h_i}} e^{\overline{\frac{(r_i - \tilde{m})^2}{2h_i(r_{i-1}, \cdots, r_0; \tilde{m}, a, b, g, h_0)}}} .$$

那么似然函数为

$$L(\tilde{m}, a, b, g, h_0) = \prod_{i=1}^{N} p_i(r_i \mid r_{i-1}, \cdots, r_0; \tilde{m}, a, b, g, h_0), \qquad (11.22)$$

取$(\hat{m}, \hat{a}, \hat{b}, \hat{g}, \hat{h}_0)$, 使在条件 $\tilde{m} > 0, 0 \le a, b, a + b, g < 1$, $h_0 > 0$条件下

$$L(\hat{m}, \hat{a}, \hat{b}, \hat{g}, \hat{h}_0) = \sup_{(\tilde{m}, a, b, g, h_0)} L(\tilde{m}, a, b, g, h_0),$$

则 $(\hat{m}, \hat{a}, \hat{b}, \hat{g}, \hat{h}_0)$ 就是参数的估计.

当模型参数确定了以后，如果令$g = V(1 - a - b)$，那么GARCH(1,1)模型可以改写为

$$h_n - V = a[(r_{n-1} - \tilde{m})^2 - V] + b[h_{n-1} - V] .$$

两边取期望，并利用

$$E(r_i - \tilde{m})^2 = E(E[(r_i - \tilde{m})^2 \mid r_{i-1}, \cdots, r_0]) = Eh_i ,$$

我们得到

$$E[h_n - V] = (a + b)E[h_{n-1} - V] .$$

再由$a + b < 1$可知$E[h_n - V] \to 0$，从而我们有渐近关系

$$\lim_{n\to\infty} E(r_n - \tilde{m})^2 = \lim_{n\to\infty} Eh_n = V ,$$

即 $V = \dfrac{g}{1 - (a + b)}$ 的含义是渐近方差.

7．3 GARCH（p,q）的参数估计

对于 GARCH(p,q)的模型参数估计, 只要递推地用$h_0, \cdots, h_{p-1}$表出

$$h_n \stackrel{\Delta}{=} h_n(\vec{a}, \vec{b}, \vec{g}, h_0, \cdots, h_{p-1}) = (b_1 h_{n-1} + \cdots + b_p h_{n-p}) + g_0 + g_1 e_{n-1}{}^2 + \cdots + g_q e_{n-q}{}^2$$

$$f_i(e_i \mid e_{i-1}, \cdots, e_0; \vec{a}, \vec{b}, \vec{g}, h_0, \cdots, h_{p-1}) = \frac{1}{\sqrt{2p h_i}} e^{\frac{e_i^2}{2h_i}} ,$$

$$L(\hat{a},\hat{b},\hat{g},h_0,\cdots.h_{p-1})=\prod_{i=p}^{N} f_i(e_i\mid e_{i-1},\cdots,x_0;\hat{a},\hat{b},\hat{g},h_0,\cdots,h_{p-1}),$$

用最大似然估计方法, 求此条件似然函数 L 在约束条件下的最大值位置, 便得模型参数的估计.

[注 1]　GARCH 模型还可以用于统计中的回归分析，即 GARCH 回归分析模型.

[注 2]　还有随机向量型误差的 GARCH 模型.

7. 4　SV 模型（随机条件异方差模型）

SV 模型就是随机条件异方差模型. 这种模型常用以描述证券交易市场中的观测序列. 它的数学表达为

$$e_n=u_n e^{\frac{1}{2}h_n}, \tag{11.23}$$

其中 $\{u_n\}$ 独立同分布,　且 $u_n\sim N(0,s_0^2)$, 而 h_n 满足如下的 AR(p)模型

$$h_n-(b_0+b_1h_{n-1}+\cdots+b_ph_{n-p})=h_n, \tag{11.24}$$

$$1-(b_1z+\cdots+b_pz^p)=0\text{ 在}\ |z|\le 1\text{ 无零点,}$$

$\{h_n\}$ 为独立同分布随机变量序列，服从 $N(0,s_h^2)$, 且 $\{u_n\}$ 与 $\{h_n\}$ 独立.　SV 模型也可以视为 Black-Scholes 模型的离散时间采样的一种推广.

对于 SV 模型，我们有

$$E(e_n\mid e_{n-1},e_{n-2},\cdots)=Eu_n\ E(e^{\frac{1}{2}h_n}\mid e_{n-1},e_{n-2},\cdots)=0\ .$$

从而由(11.23)得到平方误差的预报公式：

$$Var(e_n\mid e_{n-1},e_{n-2},\cdots)=E(e_n^2\mid e_{n-1},e_{n-2},\cdots)=Eu_n^2\ E(e^{h_n}\mid e_{n-1},e_{n-2},\cdots)$$
$$=s_0^2e^{b_0}Ee^{h_n}E(e^{b_1h_{n-1}}\mid e_{n-1},e_{n-2},\cdots)\cdots E(e^{b_ph_{n-p}}\mid e_{n-1},e_{n-2},\cdots). \tag{11.25}$$

[注 1] 如果令 $d_n=\ln e_n^2,\ v_n=\ln u_n^2$, 那么(11. 23)变成

$$d_n=h_n+v_n. \tag{11.23'}$$

这里假定 $\{e_n\}$ 是可以实测的, 于是它是量测变量, 而 $\{h_n\}$ 则是视为隐在后面的"状态变量", 因此 (h_n,d_n) 是取连续值的 p 阶隐 Markov 模型, 更确切地说, 是隐 AR 模型, 其需要估计的模型参数为 $s_0^2,s_h^2,b_0,\cdots,b_p$. 这种看法的难点是：取连续值的隐 Markov 模型还未有有效的算法. 也可以考虑用第 8 节中的 Kalman 滤波，或用第 14 章中的 EM 方法, 来递推地估计模型参数与隐状态 $(\hat{h}_{n-1},\hat{h}_{n-2},\cdots)$.　在(11.25)式得到 e_n^2 的预报中, 可以粗放地用 $e^{b_i\hat{h}_{n-i}}$ 来代替 $E(e^{b_ih_{n-i}}\mid e_{n-1},e_{n-2},\cdots)$.

关于 ARCH 模型及 SV 模型可以参考:

[1] Bollerslev, T and Krocker, K.R. ARCH modeling in finance, a review of the theory and empirical evidence. Journal of Econometrics 52, 1992.

[2] Harvey, A. and Ruis, E. Multivariate stochastic variance models. Review of Economics studies 61, 1994.

[注 2] 在经济和金融中，人们还应用下列模型

(1). 自回归标准差 ARCH 模型 (Autoregressive Standard Deviation ARCH, ASDARCH), 或绝对值 GARCH 模型 (Absolute value GARCH, AVGARCH):

$$e_n=a_1e_{n-1}+\cdots a_pe_{n-p}+h_{n-1}u_{n-1},$$

$$h_n=|g_0+g_1h_{n-1}\,|u_{n-1}|+\cdots+g_ph_{n-p}\,|u_{n-p}||$$

其中 u_n 是独立同分布的 $N(0,1\}$ 随机变量.

(2). 对于金融市场中的信息不对称所得到的数据, Nelson 在 1991 年引入指数 GARCH 模型(EGARCH)来处理它们. EGARCH 模型是

$$e_n=\sqrt{h_n}u_n,\ \ln h_n=b_0+b_1\ln h_1+\cdots+b_p\ln h_{n-p}+\sum_{k=1}^{q}g_k(|u_{n-k}|-E|u_{n-k}|+gu_{n-k}),$$

其中参数g 可以取正值或负值， 干扰u_n独立同分布, 其分布为广义误差分布, 具有分布密度

$$f(u)=\frac{c}{2^{\frac{c+1}{c}}\Gamma(\frac{1}{c})l}e^{-\frac{1}{2}|\frac{u}{l}|^c},(0<c\le\infty,\ l=\sqrt{\frac{2^{-\frac{2}{c}}\Gamma(\frac{1}{c})}{\Gamma(\frac{3}{c})}}). \qquad (11.26)$$

由计算可得 $E|u_n|=\dfrac{2^{\frac{1}{c}}\Gamma(\frac{2}{c})l}{\Gamma(\frac{1}{c})}$. 这里$c$是尾厚参数. 在$c=2$时, 有 $l=1$是 N(0,1). 在$c>2$时是厚尾分布(比 N(0,1)尾厚, 在$c<2$时是轻尾分布(比 N(0,1)尾薄. $c=\infty$时为$U[-\sqrt{3},\sqrt{3}]$分布. 对于金融数据常呈现的厚尾现象, 以用较大的c的广义误差分布作为干扰u_n 的分布更宜..

EGARCH 模型比 SV 模型更为复杂一些. 它对于金融市场中的信息不对称现象的描述, 比其它模型有较好的优势. 它对于干扰u_{n-k}取正或取负（在股市中， 它们分别表示利好消息或利空消息），各有不相对称的不同影响.

EGARCH 也常表示为无穷阶模型

$$h_n=e^{a_0+\sum_{j=1}^{\infty}u_j\{|h_{n-j}|-E|h_{n-j}|+gh_{n-j}\}}.$$

可以证明在条件

$$\sum u_i^2<\infty$$

成立时， 模型是渐近平稳的时间序列.

（3） ARCH-M 模型 这种模型也是 ARCH 与 GARCH 的推广，不同处是考虑了条件方差的变化对于条件期望的影响. 此模型为：在$e_{n-1},\cdots,e_{n-q}$已知的条件下, e_n 的条件分布密度为

$$N(g(h_n),h_n),$$

其中$\{h_n\}$与$\{\tilde{e}_n^2\}$($\tilde{e}_n\overset{D}{=}e_n-g(h_n)\sim N(0,h_n)$) 满足 ARCH 关系或 GARCH 关系. 最简单的情形是

$$g(h_n)=ch_n,\quad h_n=a_0+a_1\tilde{e}_n^2,\quad (a_0>0,0\le a_1<1)$$

的 ARCH(1)-M 模型，此时有

$$E\tilde{e}_n^2=\frac{a_0}{1-a_1},$$

$$Ee_n=E(\tilde{e}_n+ch_n)=E[\tilde{e}_n+c(a_0+a_1\tilde{e}_n^2)]=ca_0(1+\frac{1}{1-a_1})。$$

还可以求得e_n 的方差及各阶中心矩 Ee_n^k， 从而在原则上可以进一步估计参数.

(4) 积分 ARMA模型 (ARIMA)

$\{x_n\}$称为积分 ARIMA(p,q), 如果其差分序列$h_n(\overset{\Delta}{=}x_{n+1}-x_n)$为 ARMA(p,q).
类似的还有 IGARCH 模型 (积分 GARCH 模型)等.

8 二阶矩序列的滤波再访

8．1 线性滤波再访

有限个资料已知情形的线性滤波

设需要接收的是一个具有二阶矩的随机过程 S_t，它表示信号，但是在传输过程中被噪声 n_t (也是一个二阶矩过程)所干扰， 因而使实际接收到的是 $h_t = S_t + n_t$． 问题是要从接收到的 $\{h_t : 0 \le t \le N\}$ 尽可能地恢复 $\{S_t : 0 \le t \le N\}$．用资料 $\{h_t : 0 \le t \le N\}$ 作的最简单的加工就是其线性组合，也就是要用它们的线性组合来近似 $\{S_t : 0 \le t \le N\}$．即要找 $\hat{S}_t \overset{\Delta}{=} \sum_{i=1}^{N} c_i h_i$ 使

$$E(S_t - \hat{S}_t)^2 = \inf_{V \in L\{h_i : i \le N\}} E(S_t - V)^2 ,$$

其中 $L\{h_i : i \le N)\overset{\Delta}{=}\{\sum_{i=1}^{N} c_i h_i : \forall c_i, i \le N\}$．由线性代数可知上式等价于

$$\hat{S}_t \overset{\Delta}{=} \sum_{i=1}^{N} c_i h_i = \mathrm{Pr}oj_{L\{h_i : i \le N\}} S_t .$$

它又等价于满足下述方程组

$$(S_t - \hat{S}_t, h_j) \overset{\Delta}{=} E[(S_t - \hat{S}_t) h_j] = 0, (j \le N) .$$

从而得到 c_i 们满足的方程组

$$\sum_{i=1}^{N} c_i E(h_i h_j) = E(S_t h_j), (j \le N) . \qquad (11.\ 27)$$

\因为 $h_t = S_t + n_t$，所以只要知道 $R_{ij} = E(S_i S_j), r_{ij} = E(n_i n_j)$ 及信号与噪声的互相关 $E(S_i n_j)$，就可以从（１１．２７）求得系数 c_i 们．最简单情形是假定信号与噪声是不相关的，即 $E(S_i n_j) = 0, (\forall i, j \le N)$．这是在实际中最常见的情形．

8． 2 Kalman-Bucy 滤波

背景　　在观测信号中混有噪声干扰时,就要利用信号本身的规律去过滤掉噪声．这就是滤波问题．最简单的信号模型是 AR(p)模型

$$x_n = \sum_{i=1}^{p} a_i x_{n-i} + w_n , \qquad (11.\ 28)$$

它称为状态方程．在很多情形下 x_n 们无法直接测量，而能实测到的是由 $\{x_n : n \ge 0\}$ 在空间传播的方式与随机噪声 v_n 决定的．假定信号是通过多个通道接收的，而且到达各接收点时间与强度并不一致．这样，汇总而接收到的测量值，并不是信号 x_n，而是信号的现在值与它过去的值的不同权重的滑动和再加上随机噪声：

$$\sum_{j=0}^{p-1} h_j x_{n-j} + v_n .$$

我们把它记为 h_n，即

$$h_n = \sum_{j=0}^{p-1} h_j x_{n-j} + v_n , \qquad (11.\ 29)$$

它称为量测方程．

状态方程与量测方程就构成了 Kalman-Bucy 模型的核心．在内涵上它比如下的 Wiener 模型

$$h_n = x_n + v_n$$

更为广泛，而在应用中则更为切合实际..所谓 Kalman—Bucy 滤波，就是通过状态方程

$$x_n = \sum_{i=1}^{P} a_i x_{n-i} + w_n$$

和量测方程

$$h_n = \sum_{j=1}^{P-1} h_j x_{n-j} + v_n \text{ ,}$$

由量测来$\{h_n\}$估计$\{x_n\}$．显见，AR 模型可以看成 Kalman-Bucy 模型的特例．

Kalman-Bucy 模型 与滤波的一般形式

为了数学表述与实际应用的方便，我们需要把模型改写为矩阵形式．(11.28)和(11.29)这两个方程可以写成如下的向量形式

$$\begin{pmatrix} x_n \\ x_{n-1} \\ \vdots \\ x_{n-p+1} \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & \cdots & a_p \\ 1 & 0 & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{pmatrix} \begin{pmatrix} x_{n-1} \\ x_{n-2} \\ \vdots \\ x_{n-p} \end{pmatrix} + \begin{pmatrix} w_n \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

$$h_n = (h_0, h_1, \cdots, h_{p-1}) \begin{pmatrix} x_n \\ x_{n-1} \\ \vdots \\ x_{n-p+1} \end{pmatrix} + v_n \text{.}$$

我们可把上述两个方程推广地记为

$$\begin{cases} \vec{x}_n = F\vec{x}_{n-1} + \Gamma\vec{w}_n & (11.30) \\ \vec{y}_n = H\vec{x}_n + \vec{v}_n \text{ ,} & (11.31) \end{cases}$$

(且矩阵F 无特征值在$|z| \le 1$中)．在我们最初引入的 Kalman-Bucy 模型中，w_n, v_n, h_n 都是一维的随机变量，Γ 是 $p\times 1$矩阵，H 是$1\times p$ 矩阵．把它们写成向量的形式后，就可以推广为：$\vec{x}_n$ 为 p 维向量值随机过程，$\vec{w}_n$为r 维向量值随机过程，$\vec{y}_n$,$\vec{v}_n$ 为 m 维向量值随机过程，F 为$p\times p$ 矩阵，Γ 为 $p\times r$ 矩阵，H 为$m\times p$ 矩阵， 而$\vec{w}_n$, $\vec{v}_n$都是均值为 0 的,相互独立的白噪声列（即独立同分布列），且相互独立，其方差矩阵分别记为

$$E(\vec{w}_n \vec{w}_n^T) = Q \text{ , } E(\vec{v}_n \vec{v}_n^T) = R \text{ ,} \qquad (11.32)$$

其中右上角的T 表示矩阵的转置运算．这里$\{\vec{x}_n\}$是代表信号 (随机的)状态列，而$\{\vec{y}_n\}$是得到$\{\vec{x}_n\}$的信息的(随机的)观测列， $\{\vec{v}_n\}$是观测误差列． 以上的描述给出了 Kalman-Bucy 模型，两个方程分别称为状态方程和量测方程．这样推广了的向量模型包含了 $ARMA(p,q)$，对此只需取

$$m = q+1 \text{，} \vec{w}_n = \begin{pmatrix} e_n \\ e_{n-1} \\ \vdots \\ e_{n-q} \end{pmatrix} \text{，} \Gamma = \begin{pmatrix} b_0 & b_1 & \cdots & b_q \\ 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{pmatrix} \text{，} Q = s^2 I_m$$

即可，其中 I_m 为表示 m 阶单位矩阵.

Kalman-Bucy 建立这个模型的目的，是用量测到的 $\vec{y}_n$ 来递推地估计测量不到的状态 $\vec{x}_n$. 记状态 $\vec{x}_n$ 的估计为 $\hat{x}_n$. Kalman-Bucy 滤波就是要从观测列 $\{\vec{y}_n\}$ 出发，递推地对 $\{\vec{x}_n\}$ 作估计. 即利用对 $\vec{x}_{n-1}$ 的估计，记为 $\hat{x}_{n-1}$，及 y_n 得到 $\vec{x}_n$ 的估计 $\hat{x}_n$. 其原则是：在已知 $\hat{x}_{n-1}$ 时，选取矩阵 A,B 使方差矩阵

$$\Delta_n \overset{\Delta}{=} E([\vec{x}_n-(A\vec{x}_{n-1}+B\vec{y}_n)][\vec{x}_n-(A\vec{x}_{n-1}+B\vec{y}_n)]^T)$$ 最小，

其中的最小的含义是在非负定对称矩阵的意义下理解的(即对非负对称矩阵 C_1,C_2，如果 C_1-C_2 仍为非负定，则称为 $C_1\ge C_2$). 在确定了最佳的 A,B 以后，就取

$$\hat{x}_n = A\hat{x}_{n-1}+B\vec{y}_n$$

作为对 $\vec{x}_n$ 的估计. $\{\hat{x}_n\}$ 称为 $\{\vec{x}_n\}$ 的滤波列. 从 $\{y_n\}$ 得到滤波列 $\{\hat{x}_n\}$ 的装置，就称为 Kalman-Bucy 滤波器.

注意到 Δ_n 是 A 与 B 的“二次式”，所以可以由“配平方”方法可以求得 Δ_n 的最小值. 这个做法的思想很简单（只需小心地注意不要把矩阵乘法的次序弄错），写起来却有些繁琐，所以我们略去推导而直接写出以下的结论

定理１１．２６　对于 Kalman-Bucy 滤波模型（11.30），(11.31)，最佳线性滤波为

$$\hat{x}_n = A\hat{x}_{n-1}+B\vec{y}_n,$$

$$A=(I-BH)F,\quad B=\Pi_n H^T(H_n\Pi H^T+R)^{-1}\overset{\Delta}{=}K_n,$$

其中

$$\Pi_n = FP_{n-1}F^T+\Gamma Q\Gamma^T,\ P_n\overset{\Delta}{=}E[(\vec{x}_n-\hat{x}_n)(\vec{x}_n-\hat{x}_n)^T],$$

P_n 是滤波的方差矩阵的估计.

按照 Kalman-Bucy 滤波中的习惯，我们把 B 改记为 K_n，那么

$$\hat{x}_n=(I-K_nH)F\hat{x}_{n-1}+K_n\vec{y}_n = F\hat{x}_{n-1}+K_n(\vec{y}_n-HF\hat{x}_{n-1}). \qquad (11.33)$$

所以 K_n 称为（噪声）增益系数，它是量测方程所作的贡献所产生的系数.

[注]　Kalman-Bucy 模型可以是非时齐的，　即状态方程与量测方程不是常系数，而是依赖于时间 n 的情形.　此时的滤波公式也只需在（１１．３３）作相应的改变即可.　而对于如下的一维非时齐 Kalman－Bucy 模型

$$X_{n+1}=a_nX_n+b_nw_n,$$
$$Y_n=c_nX_n+d_nv_n,$$

其 Kalman-Bucy 滤波公式则可以化为

$$\hat{X}_n = a_{n-1}\hat{X}_{n-1}+\frac{c_{nn}^2}{d_n^2}(Y_n-c_na_{n-1}\hat{X}_{n-1}),$$

其中

$$s_n^2=[(a_{n-1}^2s_{n-1}^2+b_{n-1}^2)^{-1}+\frac{c_n^2}{d_n^2}]^{-1}.$$

而 $Z_n\overset{\Delta}{=}Y_n-c_na_{n-1}\hat{X}_{n-1}$ ($Z_0=Y_0$) 则称为新息过程，它满足

$$E(Z_nZ_m)=0, E(Z_nY_m)=0 (m<n).$$

以上的 Kalman-Bucy 滤波的主要构思即是，令

$$\hat{X}_n = a_{n-1}\hat{X}_{n-1} + l_n Z_n .$$

再归纳地确定待定的系数 l_n．

由上面的介绍可见，Kalman-Bucy 滤波包容了 ARMA 模型的滤波．又 ARMA 模型的预报，也可以作为 Kalman-Bucy 滤波的特殊情形，它对应与$m=1$， $H=1$， $v_n=0\ (R=0)$的情形．

然而，在实际操作时，需要知道模型参数F, H, Γ, Q, R 与初值 $P_0, \hat{x}_0$．在此， $\vec{x}_0$ 可由我们对系统的认识来给定．有时甚至可以取$\vec{x}_0=0$． P_0的取法并不十分重要，例如，可以取初值$\vec{x}_0$的方差矩阵为P_0．因为在 Kalman-Bucy 滤波的理论中有一个重要定理：在F, H Γ满足一个相当弱的条件(称为可观测条件)下，当$n\to\infty$时， P_n有一个极限P，而且后者与P_0的取法无关．这说明只要时间充分长，特殊取的P_0所产生的影响是可以消除的.同时，由于$P_n\to P$， K_n也就应该趋向某个固定的K．也就是递推进行到适当步数后， K就可以不必再变动了，这就会减少很多计算量．

综上可以看出，在操作 Kalman-Bucy 滤波前，先必须有一个模型，或者先要拟合模型的参数．为此应该采集训练 (即学习) 模型参数F, H Γ，Q, R的一些数据，先用这些训练数据来估计这些参数，这样就拟合了模型．在非时齐模型情形，模型参数就不可能由训练得到，只能由物理建模得到．也可以用后面第 15 章中的 EM 算法的思想，来轮番地更改估计与滤波，最后达到同时得到模型参数与滤波的增益系数的目的．

*9． 二阶自相似时间序列与长程相关性

9．1　统计自相似性

定义１１．２６　(自相似过程)

随机过程$\{x_t\}$称为指数为 H 的(统计)自相似过程，如果存在$H>0$，使对于任意$a>0$，过程$\{a^{-H}x_{at}: t\ge 0\}$ 的有限维分布族不依赖于参数a．

例１１．２７　(稳定过程)

时齐的独立增量过程x_t中，有很特殊的一类，称为a 阶稳定过程 $((0<a\le 2)$，其特征函数有以下形式

$$Ee^{il x_t} = e^{-ty(l)}, \qquad (１１．３４)$$

其中

$$y(l) = (-\frac{1}{s^2} + i\frac{l}{|l|}b)|l|^a,\ (0<a<2),\ 或\ y(l) = -\frac{1}{s^2}l^2\ (相当于\ a=2)．$$

由定义易见，a阶稳定过程是$H=\dfrac{1}{a}$的统计自相似过程．a阶稳定过程都是有分布密度的随机过程，但是除了$a=\dfrac{1}{2},1,2$等特殊情形以外，它们的分布密度一般都较为复杂，使用中也不很方便．

如果$1\le a\le 2, b=0, s^2=1$，则称为a 阶(标准)对称稳定过程． 此时有

$$Ee^{i l x_t} = e^{-t|l|^a} . \qquad (11.35)$$

对称稳定过程有 Markov 性, 其条件分布函数 $P(x_{t+s} \le y \mid x_s = x)$ (称为转移(分布)函数) 具有如下的条件分布密度 (称为转移密度)

$$p(t,x,y)=\frac{1}{(2p)^{\frac{1}{2}}}\int_0^\infty e^{-tu^{\frac{a}{2}}} u^{\frac{1}{2}} J_{-\frac{1}{2}}(|y-x|u)du . \qquad (11.36)$$

(在 d 维情形时

$$p(t,x,y)=\frac{1}{(2p)^{\frac{d}{2}}}\int_0^\infty e^{-tu^{\frac{a}{2}}} u^{\frac{d}{2}} J_{\frac{d}{2}-1}(|y-x|u)du ,$$

其中 $J_p(z)$ 为 p 阶 Bessel 函数:

$$J_p(z)=\sum_{k=0}^{\infty}\frac{(-1)^k}{\Gamma(k+1)\Gamma(k+p+1)}(\frac{z}{2})^{p+2k} = \int_{S^{d-1}} e^{i|z|\cos J} dJ ,$$

$\Gamma(u)=\int_0^\infty t^{u-1}e^{-t}dt$, S^m 为 m 维球面).

2 阶对称稳定过程就是 Brown运动, 1 阶对称稳定过程也称为 **Cauchy** 过程. Cauchy 过程的转移密度为

$$p(t,x,y)=\frac{s \cdot t}{p[(y-x)^2+s^2t^2]} ,$$

其数学期望为∞. 在第 3 章中, 从 0 点出发的 Brown 首次达到$x(>0)$ 的时刻t_x, 是一族依赖于 x 的随机变量, 把x改写为t后, 就可以证明$\{t_t : t \ge 0\}$是一个$\frac{1}{2}$阶(非对称的)稳定过程.

例11.28 (分数 Brown 运动 FBM 与分数 Brown场)

Gauss 过程$\{x_t : t \ge 0\}$称为指数为 a 的分数 Brown 运动 (FBM, Fractional Brownian Motion), $(0<a<2)$, 如果 $x_0=0$, $Ex_t=0$, 且

$$E(x_s x_t)=\frac{1}{2}(s^a+t^a-|s-t|^a) . \qquad (11.37)$$

容易检查它是指数为$H=\frac{a}{2}$的自相似过程. 当$a=1$时, 右边等于

$$\frac{s+t-|s-t|}{2}=\min(s,t) .$$

这说明指数为 1 的分数 Brown 运动就是 Brown 运动.

[注] 物理学家有兴趣于指数为a 的d 维分数 Brown 场，(或Levy的 Brown 场)，它是带有d 维位置参数的 Gauss 系$\{x_{\vec{x}} : \vec{x} \in R^d\}$, ($0 < a < 2$)，满足 $x_{\vec{0}} = 0$，$Ex_{\vec{x}} = 0$，且

$$E(x_{\vec{x}} x_{\vec{y}}) = \frac{1}{2}(|\vec{x}|^a + |\vec{y}|^a - |\vec{x} - \vec{y}|^a).$$

若把$x_{\vec{x}}$ 视为在地点 $\vec{x}(\in R^d)$上的随机载荷，则两点之间的载荷差的统计规律只依赖与此两点间的距离:

$$x_{\vec{x}} - x_{\vec{y}} \sim N(0, |\vec{x} - \vec{y}|^a).$$

这称为空间齐次性. 当$a = 1$且$t \geq 0$ 时，它就是指数为 1 的分数 Brown 运动，即普通的 Brown 运动.

9．2　二阶自相似性

定义11.29　宽平稳过程$\{x_t\}$称为二阶(统计)自相似的，如果存在$H > 0$，对于任意$a > 0$，有

$$a^{-H} Ex_{at} = Ex_t,\ a^{-2H} E(x_{as} x_{at}) = E(x_s x_t) . \qquad (11.38)$$

H 称为二阶自相似指数 (或 Hurst 指数).

平稳的自相似过程是二阶自相似的. 具有二阶矩的平稳过程，只要是自相似过程，就是二阶自相似过程.

定义11．30 数学期望为 0，且具有谱密度$f(l)$的宽平稳序列$\{x_t\}$，称为渐近二阶自相似的，如果在k 充分大时，其相关函数满足

$$R(mk) \sim m^{2H} R(k) . \qquad (11.39)$$

例11．31 (分数 Brown 运动的离散采样列的差分列－分数 Gauss 噪声序列(FGN))

设$\{x_t : t \geq 0\}$是指数为a 的分数 Brown 运动，$h_n = x_{n+1} - x_n$，我们来证明$\{h_n : n \geq 1\}$是二阶自相似序列. 事实上，其自相似性继承自$\{x_t : t \geq 0\}$的自相似性. 再证它是宽平稳的：对于任意n有

$$Eh_n^2 = E[(x_{n+1} - x_n)(x_{n+1} - x_n)] = E(x_{n+1}^2) + E(x_n^2) - 2E(x_{n+1} x_n)$$
$$= (n+1)^a + n^a - [(n+1)^a + n^a - 1] = 1;$$

又在$k \geq 1$时有

$$E(h_{n+k} h_n) = E[(x_{n+k+1} - x_{n+k})(x_{n+1} - x_n)]$$
$$= E(x_{n+k+1} x_{n+1}) + E(x_{n+k} x_n) - E(x_{n+k+1} x_n) - E(x_{n+k} x_{n+1})$$
$$= \frac{1}{2}[(n+k+1)^a + (n+1)^a - k^a] + \frac{1}{2}[(n+k)^a + n^a - k^a]$$

$$-\frac{1}{2}[(n+k+1)^a+n^a-(k+1)^a]-\frac{1}{2}[(n+k)^a+(n+1)^a-(k-1)^a]$$

$$=\frac{1}{2}[(k+1)^a+(k-1)^a-2k^a]\,. \qquad (11.40)$$

可见$\{h_n: n\ge 1\}$**是二阶自相似序列, 称为分数 Gauss 噪声序列** (Fractional Gaussian Noise) **.**

[注] 随机过程$\{x_t\}$**称为广义(统计)自相似的, 如果对于任意** $l>0$, **存在**$c_l>0, b_l$, **使过程**$\{c_l x_{lt}+b_l: t\ge 0\}$**与过程**$\{x_t\}$**有相同的有限维分布族 此时**$c_l, b_l$ **分别称为此广义自相似过程的尺度系数与位置参数. 显见统计自相似过程是它的特例. 广义之相似过程又称为半稳定过程.**

9．3 长程相关性

定义１１．３４ 宽平稳时间序列$\{x_n\}$ **称为长程相关的, 如果其相关函数列**$\{R(n)\}, (R(n)=E(x_n x_{m+n}))$**满足：当** $n\to\infty$**时存在一个在**∞ **附近慢变的非负函数**$L(x)$**，即满足**$\lim\limits_{t\to\infty}\frac{L(tx)}{L(t)}=1 \quad (x>0)$**的函数，使**

$$R(n)\sim\begin{cases}\dfrac{L(n)}{n^b} & (0<b<1)\\ -\dfrac{L(n)}{n^b} & (1<b<2)\end{cases}. \qquad (11.41)$$

在具有渐近地正的相关函数的情形, 长程相关的直观含义为, 宽平稳随机序列的相关函数列是不可和的，即

$$\sum_k R(n)=\infty,$$

这表示当$n\to\infty$**时相关函数**$R(n)$**趋于 0 的速度非常慢，故而称之为长程相关．但是，对于具有渐近地负的相关函数的宽平稳随机序列, 长程相关性只要求其相关函数列**$R(n)$**趋于 0,这时的内涵较为复杂，我们不予细究．事实上，定义 11.34 是以下面例 11.35 的分数 Gauss 噪声序列为背景而概括出的.**

例１１．３５（分数 Gauss 噪声序列)

设$\{x_t: t\ge 0\}$ **是指数为**a **的分数** Brown **运动, 而**$V_n=s(x_{n+1}-x_n)$, **则由上面可知**$\{V_n: n\ge 1\}$**是二阶自相似过程. 由**(11.40), **当**k **充分大时, 其相关函数序列**

$$R(k)\overset{\Delta}{=}E(V_nV_{n+k})=\frac{s^2}{2}[(k+1)^a+(k-1)^a-2k^a]$$

$$=\frac{s^2}{2}k^a[(1+\frac{1}{k})^a+(1-\frac{1}{k})^a-2]\approx\frac{s^2}{2}a(a-1)k^{a-2}.$$

可见此时间序列是长程相关的条件为 $a>1$**(这是渐近正相关的情形),或**$0<a<1$**(这是渐近负相关的情形)．又由** Gauss **性, 在稳定情形(即**n **充分大时), 此随机序列的统计特征就由其方差**

s^{2} 和自相似指数 $H=\dfrac{a}{2}$ 完全确定.

有了这两个可以自由选取的参数，就可能用这种类型的二阶自相似序列，来拟合一些看起来具有自相似机理的随机数据列. 对于长程相关的随机数据列在拟合时，就要选取 $H>\dfrac{1}{2}$. 在应用领域中，人们常用频谱密度方法．非负函数 $f(l),(l\in[0,2p])$ 称为宽平稳列的谱密度，如果其相关函数列可以表为 (此时必须有 $\lim_{k\to\infty}R(k)=0$)

$$R(k)=\int_{0}^{2p}e^{ikl}f(l)dl\ . \qquad (11.42)$$

分数 Gauss 噪声序列 $\{V_n:n\geq 1\}$ 的谱密度是由 Taqqu 等人求出的，其表达式为

$$f(l)=2\sin(2pa)\Gamma(a+1)(1-\cos l)\sum_{n=-\infty}^{\infty}|2np+l|^{-a-1}$$

在应用领域中，有一部分人认为汇率的变化，股票的收益率 (将其对数的差分列，拟合为分数 Gauss 噪声序列) 具有长程相关性． 这种认知与用 Black-Scholes 模型来描述汇率的变化，或股票的收益率恰好相反，因为后者具有 Markov 性，因此就不会长程相关．其实，实际问题非常复杂，用什么模型拟合建模，完全是各显所长．用那种模型可以得到较好的近似，需要进行多方面的比较，总结经验，不断修正模型．当然也有不少的偶然性．

关于分数 Gauss 噪声序列，可以参见 W. Leland, M. Taqqu, W. Willinger, D. Wilson 的文章 On the Self-similar Nature of Ethernet Traffic, IEEE/ACM Trans. Networking, Vol. 2, No. 1, 1994.

例１１．３６ (分数(积分)ARMA 模型)（FARIMA）

平稳 Gauss 序列 $\{x_n\}$ 称为分数 ARMA 模型，或 FARIMA(p,d,q)，如果它具有谱密度

$$f(l)=C\frac{|B(e^{il})|^{2}}{|e^{il}-1|^{2d}|A(e^{il})|^{2}}, \qquad (11.43)$$

其中 $-\dfrac{1}{2}<d<\dfrac{1}{2}$，多项式 $A(z)=z^{n}-a_{1}z^{n-1}-\cdots-a_{p-1}z-a_{p}$ 在 $|z|\leq 1$ 无零点，

$B(z)=z^{n}+b_{1}z^{n-1}+\cdots+b_{q-1}z+b_{q}$. 它等价于：$h_n\ (\overset{\Delta}{=}(I-B)^{d}x_n)$ 为 ARMA(p,q)， B 为向后平移运算：$Bx_n=x_{n-1}$， $(I-B)^{d}$ 为分数次差分运算：

$$(I-B)^{d}\overset{\Delta}{=}\sum_{k=0}^{\infty}\binom{d}{k}(-1)^{k}B^{k}, \qquad (11.44)$$

$\dbinom{d}{k}$ 为广义组合数：$\dbinom{d}{k}=\dfrac{d(d-1)\cdots(d-k+1)}{k!}$. 在 M.S.Taqqu, V.Teverosky, W.Willinger 的文章 Estimate for Long-Range Dependence: an Empirical Study (Fractals, Vol. 3, No. 4, 785-798, 1995) 中，指出了 FARIMA(0,d,0) 模型在 $0<d<\dfrac{1}{2}$ 时，是长程相关的宽平稳时间序列，其相关函数有

$$R(k)\sim a\,\frac{1}{k^{1-2d}}\ .$$

因此也是渐近二阶自相似的．在作 FARIMA(0,d,0)模型拟合时，一共只有C,d 两个参数．由于这个模型的谱密度较为简单，即

$$f(l)=\frac{C}{|2\sin\dfrac{l}{2}|^{2d}},$$

于是在建模时可以通过观测序列的谱密度估计来估计此两个参数．

长程相关性可以用 R/S 统计量作统计检验.

定义１１．３７　假设现有观测样本序列$\{X_k\},k=1,2,\cdots n$，记其样本均值为

$$\overline{X}_n=\frac{1}{n}\sum_{k=1}^{n}X_k,$$

样本方差为

$$S_n^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2$$

再记

$$W_k=(X_1+\cdots+X_k)-k\overline{X}_n \quad (1\le k\le n).$$

统计量

$$Q_n\overset{\Delta}{=}\frac{\max\{W_1,W_2,\cdots,W_n\}-\min\{W_1,W_2,\cdots,W_n\}}{S_n} \tag{11.45}$$

称为 R/S 统计量.

在数量经济中常见的时间序列$\{X_n\}$常存在下述极限：

$$g^2\overset{\Delta}{=}\frac{\lim\limits_{n\to\infty}E[\dfrac{1}{n}(\sum\limits_{j=1}^{n}X_j)^2]}{\lim\limits_{n\to\infty}E[\dfrac{1}{n}\sum\limits_{j=1}^{n}X_j^2]}.$$

在矩满足某些条件下,如果作

零假设：$\{X_n\}$非长程相关，

人们证明了在此零假设下，存在随机变量h，使$\dfrac{Q_n}{\sqrt{n}}$依分布收敛到gh，而且h分布函数为

$$F_h(x)=1+2\sum_{k=1}^{\infty}(1-4k^2x^2)e^{-2(kx)^2}.$$

此分布是非对称的，$Eh=\sqrt{\dfrac{p}{2}}\approx1.25$，$Eh^2=\dfrac{p^2}{6}$，$Varh\approx0.27$，而$h$取值的大部分都落在 3／4 和 2 之间．在给定的显著水平a下，用h的a截尾概率(例如，右截尾)的位置，就得到检验的临界值．此值

可由 F_h 的表达式，或通过临界值的表得到．如果否定了零假设，那就说明样本数据来自长程相关随机序列．可是g 的计算非常困难．而且 R/S 统计量在时间尺度太小或太大时估计误差明显．

在汇率等一些时间序列数据中，有人用其对数数据，在分数 Gauss 噪声序列或 FARIMA(0,d,0)的假设下，用 R/S 统计量检验非长程相关的零假设，如否定则就认为有长程相关性.

对于有长程相关的时间序列．预报与滤波都没有多大意义．例如.有少数大户操纵的且有大量散户的股票的价格，常常具有长程相关性．这时，预报它的价格走向很少有意义．

一般地，把长程相关性的构图与某种机制（例如说，心脏的某种状态或疾病）联系起来．但是，在渐近正相关的情形,用有限长度的数据来判别不可和性是十分困难的，这就说明了，用 R/S 统计量作检验时错误概率会很大．在水文学，气象学，地质学中，人们发现了很多长程相关的机制．

*10　非线性 AR 模型与二重 ARMA 模型简介

10.1　非线性 AR 模型

定义 11．38　如果对于随机序列x_n，存在函数 j 满足递推关系

$$x_n = h(x_{n-1},\cdots,x_{n-p}) + e_n, \qquad (11.46)$$

则x_n称为 p 阶非线性 AR 模型(Non Linear AR，NLAR)，其中e_n是白噪声序列：

$$Ee_n = 0, E(e_n e_m) = \begin{cases} 1 & (n = m) \\ 0 & (n \neq m) \end{cases}.$$

如果 e_n 还是独立同分布的随机序列，则 $h_n \overset{\Delta}{=} \begin{pmatrix} x_n \\ x_{n-1} \\ \vdots \\ x_{n-p+1} \end{pmatrix}$ 是时间离散状态连续的 Markov 链．

对于模型（11．46），有下面简洁的收敛定理：

定理 11．39　（安－黄, 1996）　设 e_n 独立同分布，有正的分布密度 $Ee_n = 0$, $Ee_n^2 = 1$，e_n 与 $\{x_{n-1},\cdots,x_0\}$ 独立, 函数 h 满足

$$h(y) = \sum_{i=1}^{p} a_i y_i + o(\| y \|) \qquad (y = (y_1,\cdots,y_p), \| y \| > R) \qquad (11.47)$$

$(\| y \| = \sqrt{y_1^2 + \cdots + y_p^2})$, 且 $h(y) - \sum_i a_i y_i$ 在任意有界集上有界, 又多项式 $A(z) = 1 - a_1 z - \cdots - a_{p-1} z^{p-1} - a_p z^p$ 在$| z | \leq 1$ 无零点. 那么, 对于下述非线性 AR 模型

$$x_n = h(x_{n-1},\cdots,x_{n-p}) + e_n$$

所定义的Markov 链的 n 步转移函数 $P^{(n)}(x,\Lambda)=P(x_n\in\Lambda\mid x_0=x)$，存在分布函数 $F(x)$ 及 $0<r<1$ 使任意 Borel 集合 Λ，均有

$$\frac{1}{r^n}\mid p^{(n)}(x,\Lambda)-F(\Lambda)\mid\to 0\,. \qquad (11.48)$$

［注］　事实上，这里能得到的收敛性要更强得多，即谓全变差收敛，并称此 Markov 链是几何遍历的．我们在本书中不再解释全变差收敛的概念．

还有一个很容易验证的充分条件，这就是下面的定理

定理１１．４０　将定理１１．39 中的关于函数 h 的条件改为：
存在 $0<r<1, C>0$，使

$$\mid h(x_1,\cdots,x_p)\mid\le r\max_i\mid x_i\mid+C \qquad (11.49)$$

（或存在 $R>0$ 使 $\mid h(x_1,\cdots,x_p)\mid\le r\max_i\mid x_i\mid+C\quad(\|x\|>R)$），则（１１．４８）成立．

作为定理１１．39 或者定理１１．40 的直接推论，我们有

推论１１．41（Bhattachaya-Lee，1995）　设 e_n 独立同分布，有正的分布密度，$Ee_n=0, Ee_n^2=1$，e_n 与 $\{x_{n-1},\cdots,x_0\}$ 独立, h 为局部收缩型线性增长：存在 $R,a_1,\cdots,a_p>0$，满足

$$\mid h(x_1,\cdots,x_p)\mid\le\sum_{i=1}^{p}a_i\mid x_i\mid+C,(\sum_i a_i<1, x=(x_1,\cdots,x_p),\|x\|>R). \qquad (11.50)$$

那么, 下述非线性 AR 模型

$$x_n=h(x_{n-1},\cdots,x_{n-p})+e_n$$

所定义的Markov 链的 n 步转移函数 $P^{(n)}(x,\Lambda)=P(x_n\in\Lambda\mid x_0=x)$，存在分布函数 $F(x)$ 及 $0<r<1$ 使任意 Borel 集合 Λ，均有

$$\frac{1}{r^n}\mid p^{(n)}(x,\Lambda)-F(\Lambda)\mid\to 0\,.$$

推论１１．４２　若 e_n 同定理１１．39，$h(x_1,\cdots,x_p)$ 有界，则（１１．４７）成立．

对于几何遍历的Markov 链 x_n，记其概率不变分布为 $F(x)$，于是有 Markov 链 x_n 的遍历定理，即不管 Markov 链 x_n 的初始值是什么，对于任意有界 Borel 函数 f，恒有

$$P(\lim_{N\to\infty}\frac{1}{N}(f(x_{p+1})+\cdots+f(x_{p+N}))=\int f(x)F(dx))=1.$$

特别地，对于任意 Borel 集 Λ，有

$$P(\lim_{N\to\infty}\frac{1}{N}(I_\Lambda(x_{p+1})+\cdots+I_\Lambda(x_{p+N}))=\int_\Lambda F(dx))=1, \qquad (11.51)$$

即在系统达到稳定以时，Markov 链落在集合 Λ 的概率近似地等于它在沿轨道出现的频率．

例１１．４３　（门限 AR 模型，TAR）

设有延迟参数 d ：$l\le d\le p$，及门限参数 $-\infty=r_1<\cdots<r_l<r_{l+1}=\infty$，而

$$x_n=\sum_{i=1}^{l}(a_{i,0}+a_{i,1}x_{n-1}+\cdots+a_{i,p}x_{n-p+1})I_{(r_k,r_{k+}]}(x_{n-d})+e_n\,. \qquad (11.52)$$

其中 e_n 独立同分布，有正的分布密度，$Ee_n=0, Ee_n^2=1$，e_n 与 $\{x_{n-1},\cdots,x_0\}$ 独立．那么，只要

$$r = \max_i \sum_{j=1}^{p} |a_{ij}| < 1,$$

定理１１．４１的条件就满足，从而 Markov 链x_n是几何遍历的.

10．２　非线性 AR 模型的常见例子

用非线性 AR 模型建模，大体可以出现以下数种情形

１．参数模型；例如

（1）分片线性 AR 模型：

$$x_n = \sum_{i=1}^{l} (a_{i,0} + a_{i,1}x_{n-1} + \cdots + a_{i,p}x_{n-p+1}) I_{\Delta_i}(x_{n-1}, \cdots, x_{n-p+1}) + e_n. \quad (１１．５３)$$

（空间 R^p 分为 m 个两两不交的区域 $\Lambda_1, \cdots, \Lambda_m$，其中 Λ_1 无界，而 $\Lambda_2, \cdots, \Lambda_m$ 都有界）．模型参数为 $a_{i,k} \quad (i \le l, 0 \le k \le p)$．

（2）拟线性 AR 模型：

$$x_n = a_0 + \sum_{i=1}^{l} a_i f_i(x_{n-1}, \cdots, x_{n-p+1}) + e_n,$$

其中 $f_i \quad (i \le l)$ 是已知函数，模型参数是 $a_0, \cdots, a_p$．

（3）指数 AR 模型：

$$x_n = a_0 + \sum_{k=1}^{p} (a_k + b_k e^{-g x_{n-k}^2}) x_{n-k} + e_n.$$

模型参数是 $a_0, \cdots, a_p, b_1, \cdots, b_p$．

（4）分式 AR 模型：

$$x_n = \frac{P(x_{n-1}, \cdots, x_{n-p})}{Q(x_{n-1}, \cdots, x_{n-p})} + e_n.$$

模型参数是分子与分母两个多项式的系数.

（5）双线性 ARMA 模型：

$$x_n = \sum_{k=1}^{p} a_k x_{n-k} + \sum_{j=1}^{q} b_j e_{n-j} + \sum_{k=1}^{p} \sum_{j=1}^{q} c_{kj} x_{n-k} e_{n-j}.$$

对于参数模型，$h_n \stackrel{\Delta}{=} (x_n, \cdots, x_{n-p+1})$ 是 Markov 链．首先要加条件使它有不变分布．然后，在此条件下估计模型参数.

２．非参数模型　一般的模型如（１１．４６），其中 h 是未知的待估函数，可以参照统计中的核估计方法，找出估计它的方法作为建模（但是还需要保证 Markov 链有遍历性）．

模型（１１．４６）有以下的几种向随机条件方差的推广（其中也包括了一些半参数模型）：

（１）非线性 ARCH 模型：

$$x_n - j(x_{n-1},\cdots,x_{n-p}) = e_n \text{，} \quad e_n = u_n\sqrt{h(e_{n-1},\cdots,e_{n-q})} \text{，} \qquad u_n \text{ 是白噪声．}$$

（２）非线性 GARCH 模型：

$$x_n - j(x_{n-1},\cdots,x_{n-p}) = e_n \text{，} \quad e_n = u_n\sqrt{h_n} \text{，}$$

其中 $h_n = g(e_{n-1},\cdots,e_{n-q},h_{n-1},\cdots,h_{n-r})$，$j\,,g$ 是待估的模型函数．最常见的情形是辅以以下的参数模型（即成为半参数模型）

$$h_n = a_0 + a_1 e_{n-1}^2 + \cdots + a_q e_{n-q}^2 + b_1 h_{t-1} + \cdots + b_r h_{n-r} \quad (a_i, b_j \ge 0, 0 \le i \le q, j \le r) \text{．}$$

可以证明此种非线性 GARCH 模型的几何遍历的条件为 $a_1 + \cdots + a_q < 1$（参见［A］）．有时觉得此条件的限制过大，相应地有下面的一类半参数模型

（３）非线性 $b-$GARCH 模型：

$$x_n - j(x_{n-1},\cdots,x_{n-p}) = e_n \text{，} \quad e_n = u_n\sqrt{h_n} \text{，}$$

$$h_n = a_0 + a_1 e_{n-1}^{2b_1} + \cdots + a_q e_{n-q}^{2b_1} + b_1 h_{n-1}^{2b_2} + \cdots + b_r h_{n-r}^{2b_2}$$

$(a_i, b_j \ge 0, 0 \le i \le q, j \le r, 0 \le b_1 < 1, 0 \le b_2 \le 1)$．非线性 $b-$ARCH 模型类似．

（４）函数型随机条件方差非线性 AR 模型：

$$x_n = j(x_{n-1},\cdots,x_{n-p}) + u_n S(x_{n-1},\cdots,x_{n-p}) \qquad (u_n \text{ 是白噪声）．}$$

常见的非参数模型还有

（５）可加 AR 模型：

$$x_n = a + \sum_{k=1}^{p} h_k(x_{n-k}) + e_n \text{，}$$

其中 $h_k \quad (k \le p)$ 是未知的待估函数．

（６）函数系数 AR 模型：

$$x_n = a + \sum_{k=1}^{p} h_k(x_{n-k})x_{n-k} + e_n.$$

３．半参数模型：一般形式为

$$x_n = h(x_{n-1}, \cdots, x_{n-p}, J) + e_n.$$

除了２．中的（３）是半参数模型以外，还有，例如

（１）广义线性 AR 模型：

$$x_n = h(\sum_{k=1}^{p} a_k x_{n-k}) + e_n,$$

这里待估的既有未知参数$a_1, \cdots, a_p$，又有未知函数h．

（２）加法半参数 AR 模型：

$$x_n = a_0 + \sum_{k=1}^{p} a_k x_{n-k} + h(x_{n-1}, \cdots, x_{n-p}) + e_n.$$

（３）逆 AR 模型：

$$x_n = h(\sum_{k=1}^{p} a_k x_{n-k} + e_n).$$

关于随机条件方差模型的几何遍历性，有以下定理

定理１１．４４ 对于函数型随机条件方差非线性 AR 模型

$$x_n = j(x_{n-1}, \cdots, x_{n-p}) + u_n S(x_{n-1}, \cdots, x_{n-p}) \qquad (１１．５４)$$

满足

（１） u_n独立同分布，有正的分布密度，$Eu_n = 0, Eu_n^2 = 1$，u_n与$\{x_{n-1}, \cdots, x_0\}$独立；

（２） $j(y) = \sum_{i=1}^{p} a_i y_i + o(\| y \|) \qquad (y = (y_1, \cdots, y_p), \| y \| > R)$

且$j(y) - \sum_i a_i y_i$在任意有界集上有界，$A(z) = 1 - a_1 z - \cdots - a_{p-1} z^{p-1} - a_p z^p$ 在$|z| \le 1$ 无零点；

（３）在任意有界集上$S(x_1, \cdots, x_p)$有上界与非０的下界，而且

$$S(x) \overset{\Delta}{=} S(x_1, \cdots, x_p) = o(\| x \|) \qquad (x \to \infty), \qquad (１１．５５)$$

那么x_n是几何遍历的.

例１１．45　作为定理１１　．44的直接推论，b — ARCH模型

$$e_n = u_n\sqrt{h_n}\text{，} h_n = a_0 + a_1 e_{n-1}^{2b} + \cdots + a_p e_{n-p}^{2b}$$

$(a_i \geq 0, 0 \leq i \leq p, 0 \leq b < 1)$是几何遍历的.

［注］　对于ARCH模型$e_n = u_n\sqrt{h_n}$，$h_n = a_0 + a_1 e_{n-1}^2 + \cdots + a_p e_{n-p}^2$

$(a_i \geq 0, 0 \leq i \leq p)$，定理 11.44 的方法并不适用．但是可以利用 Markov 链中著名的 Tweedie 定理（相当于常微分方程中的用 Lyapunov 函数得到收敛性的随机版本），可以直接证明：当$a_1 + \cdots + a_p < 1$时e_n是几何遍历的．这正是本章第6节中作参数估计时所要求的前提．

10．３　二重ARMA模型

如果ARMA模型中的参数序列不是常数，而是随机变量序列，并且也是一个 ARMA 模型，则就成为二重 ARMA模型，也称ARMA-ARMA模型．　一般地，用这种模型建模过于复杂，所以在应用中常见的是 AR(1)-AR(1)，AR(1)-MA(q)等较为简单的情形：

AR(1)-AR(1)模型是

$$x_n = a_0 + (a_1 + h_n)x_{n-1} + e_n\text{，} h_n = bh_{n-1} + v_n\text{，}$$

其中$\{e_n\},\{v_n\}$是相互独立的独立同分布随机序列．这种模型的拟合，需要估计５个参数：$a_0, a_1, b, Var(e_n), Var(v_n)$．

AR(1)-MA(q)模型是

$$x_n = a_0 + (a_1 + h_n)x_{n-1} + e_n\text{，} h_n = v_n + b_1 v_{n-1} + \cdots + b_q v_{n-q}\text{，}$$

其中$\{e_n\},\{v_n\}$是相互独立的独立同分布随机序列．这种模型的拟合，需要估计的参数为：$a_0, a_1, b_1, \cdots, b_q, Var(e_n), Var(v_n)$．

对于 AR(1)－MA(1)模型的 Markov 链x_n具有不变分布的条件可参考 ［A］．

习题　１１

１．设x_t是轨道为t的连续函数的Gauss过程，$Ex_t = m(t), Cov(x_s, x_t) = R(s,t)$，$G(t)$是单调递增函数．如果存在随机变量$h_t$，使　$Var(h_t) < \infty$，而且当$\max_i (t_{i+1}^{(n)} - t_i^{(n)}) \xrightarrow{n\to\infty} 0$时有

$$E\left|\sum_i x_{t_i^{(n)}}[G(t_{i+1}{}^{(n)}) - (G(t_i{}^{(n)})] - h_t\right|^2 \to 0\text{，}$$

这里$\{t_i{}^{(n)}\}$是半开区间$(0,t]$的一个划分．那么，定义　$\int_0^t x_s dG(s) = h_t$．证明h_t也是Gauss过程．再求$Eh_t, Cov(h_s, h_t)$．

２．设x_t, h_t是相互独立的Gauss过程，证明$x_t + h_t$也是Gauss过程.

3．设x_t, h_t是宽平稳过程，而且平稳相关，即$Cov(x_s, h_t)$只与$t-s$有关．证明x_t+h_t也是宽平稳过程．

4．设x_t是轨道为t的连续函数的Gauss过程，$Ex_t = m(t), Cov(x_s, x_t) = R(s,t)$，$G(t)$是任意单调递增函数．Gauss过程$x_t$的特征泛函定义为

$$\Phi_x(G) = Ee^{i\int_0^t x_s dG(s)}.$$

由此特征泛函的显式表示，对于$t_1 < t_2 < t$，取$G(t) = J_1 \cdot I_{[0,t_1]}(t) + J_2 \cdot I_{(0,t_2]}(t)$，用以得到$(x_{t_1}, x_{t_2})$的联合特征函数$j_{x_1,x_2}(J_1, J_2) = Ee^{i(J_1x_1+J_2x_2)}$的显式表示．

5．设x_t如题1, 对于$t_1 < t_2 < t_3 < t_4 < t$，求$E(x_{t_1}x_{t_2}x_{t_3}x_{t_4})$．

6．设h_1, h_2, J_1, J_2是相互独立的随机变量．而$x_t = h_1 Cos(l_1 t + J_1) + h_2 Cos(l_2 t + J_2)$，$(l_1 \neq l_2)$．问$x_t$是否是平稳过程　？什么时候是宽平稳？　推广此结论.

7．复随机过程$x_t = h_1 e^{il_1 t} + h_2 e^{il_2 t}$，（$l_1 \neq l_2$，$h_1, h_2$相互独立）．问$x_t$是否是宽平稳？　推广此结论.

8．设x_t是平稳过程，问$h_t = x_s x_t$是否为平稳过程　？

9．对于AR模型$x_n - a_1 x_{n-1} - \cdots - a_p x_{n-p} = e_n$, 证明它有形式为$f(l) = C\dfrac{1}{|A(e^{il})|^2}$的谱密度，其中多项式$A(z) = 1 - a_1 z - \cdots - a_{p-1} z^{p-1} - a_p z^p$在$|z| \le 1$内无零点.

10．　对于MA模型$x_n = b_0 e_n + b_1 e_{n-1} + \cdots + b_q e_{n-q}$, 证明它有形式为$f(l) = C|B(e^{il})|^2$的谱密度.

11．对于ARMA模型$x_n - a_1 x_{n-1} - \cdots - a_p x_{n-p} = b_0 e_n + b_1 e_{n-1} + \cdots + b_q e_{n-q}$，其中多项式$A(z) = 1 - a_1 z - \cdots - a_{p-1} z^{p-1} - a_p z^p$在$|z| \le 1$无零点．证明它有形式为

$f(l) = C\dfrac{|B(e^{il})|^2}{|A(e^{il})|^2}$的谱密度.

12．设e_n独立同分布，有正的分布密度，$Ee_n = 0, Ee_n^2 = 1$，e_n与$\{x_{n-1}, \cdots, x_0\}$独立．　函数h在任意有界集上有界，且$h(x) = o(\|x\|)$　　$(\|x\| \to \infty)$．证明由

$$x_n = a_0 + a_1 x_{n-1} +, \cdots, + a_p x_{n-p} + h(x_{n-1}, \cdots, x_{n-p}) + e_n$$

确定的Markov链在$A(z) = 1 - a_1 z - \cdots - a_{p-1} z^{p-1} - a_p z^p$在$|z| \le 1$内无零点时为几何遍历.

13．给出分片线性AR模型$x_n = \sum_{i=1}^{l} (a_{i,0} + a_{i,1} x_{n-1} + \cdots + a_{i,p} x_{n-p+1}) I_{\Delta_i}(x_{n-1}, \cdots, x_{n-p+1}) + e_n$

（设e_n同题12．而空间R^p分为m个两两不交的区域$\Lambda_1, \cdots, \Lambda_m$，其中$\Lambda_1$无界，而$\Lambda_2, \cdots, \Lambda_m$都有界）几何遍历的条件（用模型参数为$a_{i,k}$　　$(i \le l, 0 \le k \le p)$描述）.

龚光鲁，钱敏平著　应用随机过程教程 – 与在算法和智能计算中的应用
清华大学出版社，2003

第 12 章　连续时间连续状态的 Markov 过程，鞅，Ito 积分与随机微分方程

1 ．连续时间连续状态的 Markov 过程

1．1　平稳 Gauss 过程

定义１２．１　Gauss 平稳过程{ ξ_t ,t∈R} 称为 Markov 型的，如果其相关函数

$B(s)\overset{\Delta}{=}E(x_t x_{t+s})$ 是连续函数，而且对任意 s≥0,t∈R 有%

$$\inf_{n,i\le n,t_i\le t,a_i\in R} E|x_{t+s}-\sum_{i\le n} a_i x_{t_i}|^2=\inf_{a\in R} E|x_{t+s}-a x_t|^2 . \qquad (12．1)$$

其含义是，在均方距离$\|x\|=\sqrt{Ex^2}$ 下，若用过程在时刻t以前的资料$\{x_u : u\le t\}$的有限线性组合去近似x_{t+s}，其最佳的近似只需在 $\overline{L}(x_{[0,t]})\overset{D}{=}\{a x_t : a$为任意实数$\}$中寻找。

对于 Markov 型的 Gauss 平稳过程，(12 .1)右方必然在某个(依赖于 s 的) a 上达到，我们记此a 为$a(s)$．又由于(12.1)成立，那么，对任意$u\ge 0$及 l_1， l_2有

$$E|x_{t+s}-(l_1 x_{t-u}+l_2 x_t)|^2\ge E|x_{t+s}-a(s)x_t|^2 .$$

从而，对任意$u,s>0$，由投影公式有

$$E[(x_{t+s}-a(s)x_t)x_{t-u}]=0 ,$$

即

$$B(s+u)=a(s)B(u) . \qquad (12．2)$$

取$u=0$，我们得到　$a(s)=\dfrac{B(s)}{B(0)}$．在(12.2)两端除以$B(0)$后便得到

$$a(s+u)=a(s)a(u) \qquad (s,u\ge 0) .$$

再由$B(s)$的连续性，即$a(s)$是连续函数，及$a(s)\le 1=a(0)$推出

$$a(s)=B(0)e^{-b\cdot t} \qquad (b,s\ge 0) .$$

记$g=B(0)$，便得　B(s)= g　e^{-bs}．$b,g>0$的情形称为非退化情形．当$s<0$时，由平稳性我们显见有$B(s)=B(-s)$，合起来便成为

$$B(s)=g\ e^{-b|s|} . \qquad (12．3)$$

此条件也是具有连续相关函数 $B(t)$ 的实值非退化 Gauss 平稳过程 $\{x_t:-\infty<t<\infty\}$ 是 Markov 型的充要条件.

定理１２．２ 非退化的 Markov 型的 Gauss 平稳过程$\{x_t:-\infty<t<\infty\}$，在任意有限个时刻所得的随机向量$(x_{t_1},\cdots,x_{t_n})$的联合分布具有密度.

证明 不失一般性，我们可以假定 $Ex_t\equiv 0$，$g=b=1$，$t_1<\cdots<t_n$. 由(12．1)及(12．2)我们得到

$$E[(x_{t_n}-e^{-(t_n-t_{n-1})}x_{t_{n-1}})x_{t_j}]=0,(j\le n-1).$$

利用$\{x_t:-\infty<t<\infty\}$的 Gauss 性就得到$(x_{t_n}-e^{-(t_n-t_{n-1})}x_{t_{n-1}})$与$(x_{t_{n-1}},\cdots,x_{t_1})$相互独立.

下面我们用数学归纳法证明$(x_{t_n},\cdots,x_{t_1})$有联合密度。

当$n=1$时，由于 $x_{t_1}\sim N(0,1)$，显见它有密度．今作归纳法假定，设$n=m-1$时已有密度 $r_{(t_1,\cdots,t_{m-1})}(x_1,\cdots,x_{m-1})$．往证在$n=m$时结论也成立．为此我们记

$$h_m\stackrel{D}{=}x_{t_m}-e^{-(t_m-t_{m-1})}x_{t_{m-1}}.$$

那么它的方差为

$$Var(h_m)=E(x_{t_n}-e^{-(t_n-t_{n-1})}x_{t_{n-1}})^2$$

$$=B(0)-2e^{-(t_m-t_{m-1})}B(t_m-t_{m-1})+e^{-2(t_m-t_{m-1})}B(0)=1-e^{-2(t_m-t_{m-1})}.$$

即 $h_m\sim N(0,1-e^{-2(t_m-t_{m-1})})$). 由于它与$(x_{t_{m-1}},\cdots,x_{t_1})$独立，从而$(h_m,x_{t_{m-1}},\cdots,x_{t_1})$有密度:

$$p_{h_m,x_{t_1},\cdots,x_{t_{m-1}}}(y,x_1,\cdots,x_{m-1})$$

$$=r_{(t_1,\cdots,t_{m-1})}(x_1,\cdots,x_{m-1})\frac{1}{\sqrt{2p(1-e^{-2(t_m-t_{m-1})})}}e^{-\frac{y^2}{2(1-e^{-2(t_m-t_{m-1})})}}.$$

利用随机变量密度函数的变换公式，并利用 $y=x_m-e^{-(t_m-t_{m-1})}x_{m-1}$，就得到$(x_{t_m},\cdots,x_{t_1})$的密度表达式为

$$r_{(t_1,\cdots,t_m)}(x_1,\cdots,x_m)$$

$$= r_{(t_1,\cdots,t_{m-1})}(x_1,\cdots,x_{m-1})\frac{1}{\sqrt{2p(1-e^{-2(t_m-t_{m-1})})}}e^{-\frac{(x_m-x_{m-1}e^{-(t_m-t_{m-1})})^2}{2(1-e^{-2(t_m-t_{m-1})})}}$$

$$=\frac{1}{\sqrt{2p}}e^{-\frac{x_1^2}{2}}\prod_{k=1}^{m}\frac{1}{\sqrt{2p(1-e^{-2(t_k-t_{k-1})})}}e^{-\frac{(x_k-x_{k-1}e^{-(t_k-t_{k-1})})^2}{2(1-e^{-2(t_k-t_{k-1})})}} \qquad (12．4)$$

推论１２．３　对于相关函数为$B(s)=ge^{-b|s|}$**的 Gauss 平稳过程**$\{x_t:-\infty<t<\infty\}$**，及**$t_1<\cdots<t_n$**，**$(x_{t_n},\cdots,x_{t_1})$ **有密度**

$$r_{(t_1,\cdots,t_n)}(x_1,\cdots,x_n)$$

$$=\frac{1}{\sqrt{2pg}}e^{-\frac{x_1^2}{2}}\prod_{k=1}^{n}\frac{1}{\sqrt{2pg(1-e^{-2b(t_k-t_{k-1})})}}e^{-\frac{(x_k-x_{k-1}e^{-b(t_k-t_{k-1})})^2}{2g(1-e^{-2b(t_k-t_{k-1})})}}. \qquad (12．4)'$$

推论１２．４ 相关函数为$B(s)=ge^{-b|s|}$**的 Gauss 平稳过程**$\{x_t:-\infty<t<\infty\}$**，具有以下的 Markov 性质：对于任意**n**及任意**$t_1<\cdots<t_{n+1}$**，在给定**$(x_{t_n},\cdots,x_{t_1})=(x_n,\cdots,x_1)$**的条件下，**$x_{t_{n+1}}$**的条件分布密度只与**$x_{t_n}=x_n$**有关，而与**$(x_{t_{n-1}},\cdots,x_{t_1})=(x_{n-1},\cdots,x_1)$**无关（即在**$x_{t_n}=x_n$**已知的条件下，**$x_{t_{n+1}}$**与**$(x_{t_{n-1}},\cdots,x_{t_1})$**条件独立）.**

证明　**我们把在**$(x_{t_n},\cdots,x_{t_1})=(x_n,\cdots,x_1)$**的条件下**$x_{t_{n+1}}$**的条件分布密度记为**$p_{t_{n+1}|t_n,\cdots,t_1}(x_{n+1}\,|\,x_n,\cdots,x_1)$**，而把在**$x_{t_n}=x_n$**的条件下**$x_{t_{n+1}}$**的条件分布密度记为**$p_{t_{n+1}|t_n}(x_{n+1}\,|\,x_n)$**．由(12．4)'，我们有**

$$p_{t_{n+1}|t_n,\cdots,t_1}(x_{n+1}\,|\,x_n,\cdots,x_1)=\frac{r_{(t_{n+1},\cdots,t_1)}(x_{n+1},\cdots,x_1)}{r_{(t_n,\cdots,t_1)}(x_n,\cdots,x_1)}$$

$$=\frac{1}{\sqrt{2pg(1-e^{-2b(t_{n+1}-t_n)})}}e^{-\frac{(x_{n+1}-x_ne^{-b(t_{n+1}-t_n)})^2}{2(1-e^{-2b(t_{n+1}-t_n)})}}$$

$$=\frac{r_{(t_{n+1},t_n)}(x_{n+1},x_n)}{r_{t_n}(x_n)}=p_{t_{n+1}|t_n}(x_{n+1}\,|\,x_n).$$

再对x_{n+1}**在某个集合**Λ**上积分，便得**

$$P(x_{t_{n+1}} \in \Lambda \mid x_{t_n} = x_n, \cdots, x_{t_1} = x_1) = P(x_{t_{n+1}} \in \Lambda \mid x_{t_n} = x_n)\ .$$

这就是时间与状态都连续的随机过程$\{x_t : -\infty < t < \infty\}$的 Markov 性．因为它表明，在已知此过程的现在时刻（即 $t = t_n$ ）的取值 $x_{t_n} = x_n$ 的条件下，过程在过去的取值$(x_{t_{n-1}}, \cdots, x_{t_1}) = (x_{n-1}, \cdots, x_1)$并不影响将来（即$t + s = t_{n+1}$）的$x_{t_{n+1}}$取值的统计规律．因此这正是第 6 章中的连续时间离散状态 Markov 性在连续时间连续状态情形的推广．

推论１２．５ $(x_{t_n}, \cdots, x_{t_1})$**的联合密度可以写为：**

$$r_{(t_1, \cdots, t_n)}(x_1, \cdots, x_n) = p_{t_1}(x_1) p_{t_2|t_1}(x_2 \mid x_1) \cdots p_{t_n|t_{n-1}}(x_n \mid x_{n-1})\ .$$

它即是"起始"密度 $p_{t_1}(x_1)$ 与一系列条件密度的乘积．

以下我们只考虑$t \ge 0$，且$Ex_t \equiv 0, b, g > 0$的情形．我们把t_n改记为 s，$t_n - t_{n-1}$改记为t，再将$p_{t+s|s}(y \mid x)$（注意它的表达式与s无关）改记为 $p(t, x, y)$ ．

命题１２．６ $\{p(t, x, y) : t \ge 0, x, y \in R\}$**满足 Kolmogorov-Chapman 方程：**

$$p(t + s, x, y) = \int p(t, x, z)\ p(t, z, y) dz\ . \qquad (12．5)$$

证明　为了看起来对称，令 $t_1 + s = t_2,\ t_2 + t = t_3,\ x = x_1, z = x_2, y = x_3$**．于是**

$$\int p(s, x, z) p(t, z, y) dz = \int p_{t_2|t_1}(x_2 \mid x_1) p_{t_3|t_2, t_1}(x_3 \mid x_2, x_1) dx_2$$

$$= \int \frac{r_{t_1, t_2}(x_1, x_2)}{r_{t_1}(x_1)} \frac{r_{t_1, t_2, t_3}(x_1, x_2, x_3)}{r_{t_1, t_2}(x_1, x_2)} dx_2$$

$$= \int p_{t_3, t_2|t_1}(x_3, x_2 \mid x_1) dx_2 = p_{t_3|t_1}(x_3 \mid x_1) = p(s + t, x, y)\ .$$

１．２　时间与状态都连续的时齐 Markov 过程

定义１２．６　假定对于任意的n及n个时刻$t_1 < \cdots < t_n$，$(x_{t_n}, \cdots, x_{t_1})$都有正的联合密度 $r_{(t_1, \cdots, t_n)}(x_1, \cdots, x_n)$，且其条件分布满足：对任意的$t_1 < \cdots < t_{n+1} = t_n + t$，$t, s = t_n \ge 0$，都有不依赖$t_n$的关系：

$$p_{t_n + t|t_n, t_{n-1}, \cdots, t_1}(y \mid x, x_{n-1} \cdots, x_1) = p_{t_n + t|t_n}(y \mid x) = p_{s+t|s}(y \mid x) \text{（记成 } p(t, x, y) \text{）}, \qquad (12．6)$$

则称随机过程$\{x_t : t \ge 0\}$为时齐的 Markov 过程，并称 $p(t, x, y)$ 为此 Markov 过程的转移密

度．

推论１２．７ 时齐的 Markov 过程的转移密度满足：

(P.1) $p(t,x,y)\ge 0,\ \int\ p(t,x,y)dy=1$；

(P.2) Chapman-Kolmogorov 方程：$p(t+s,x,y)=\int\ \ p(t,x,z)\ \ p(t,z,y)dz$．

［注 1］转移密度 $p(t,x,y)$ 是转移概率 $p_{ij}(t)$ 在状态空间为实数集时的对应表示．

［注 2］时齐 Markov 过程也可以只对 t>0 时要求 $(x_{t_n},\cdots,x_{t_1})$ 都有正的联合密度．也就是说，初值 x_0 可以特殊一些，不必要求它有密度．因为我们可以利用时齐性补充定义在条件 $x_0=x$ 下，x_t 的条件密度为 $p(t,x,y)$．从而可以讨论 x_0 取离散值或有密度的情形．也就是说对于在本书中在时间区间 $(0,\infty)$ 上定义的时齐 Markov 过程 $\{x_t:t>0\}$，在加进 $t=0$ 的随机变量 x_0 以后，总可以扩展为时间区间 $[0,\infty)$ 上时齐的 Markov 过程 $\{x_t:t\ge 0\}$．

［注 3］ 定义中关于存在正的有限维联合分布密度的要求不是必须的．由于没有这个假定时需要用测度论的知识，所以在本书中只取最简单的情形．

例１２．８（OU 过程） $b,g>0$ 的 Markov 型 Gauss 平稳过程是时齐的 Markov 过程，其转移密度为

$$p(t,x,y)=\frac{1}{\sqrt{2pg\,(1-e^{-2bt})}}e^{-\frac{(y-xe^{-bt})^2}{2(1-e^{-2bt})}}. \qquad (12.7)$$

而且有

$$\lim_{t\to\infty}\ p(t,x,y)=\frac{1}{\sqrt{2pg}}e^{-\frac{y^2}{2g}}\overset{\Delta}{=}j\,(y)$$ （即极限分布为 $N(0,g)$，与 x,b 都无关）．

在 Chapman-Kolmogorov 方程中令 $s\to\infty$，并把 z 改记为 x，便得到

$$j\,(y)=\int p(t,x,y)j\,(x)dx. \qquad (12.8)$$

即 $j\,(y)$ 是一个满足（１２．８）分布密度，称为 $p(t,x,y)$ 的不变密度．它是时间连续状态离散的 Markov 链的不变分布在连续状态情形的对应表示．

注 具有(12.7)形式的转移密度的 Markov 过程，称为参数为 (b,g) 的 OU 过程（Ornstein-Uhlenbeck 过程）．

OU 过程在有限个采样时刻上的样本，可以用如下的随机模拟得到．

相关函数为 $B(t)=ge^{-b|t|}$ 的 OU 过程在等距采样点 $t_1,\cdots,t_n$ 上的样本的随机模拟，可以由下面步骤得到：

令 $y_1,\cdots,y_n$ 为独立的 $N(0,1)$ 随机数，$\Delta t = t_{k+1} - t_k$．定义

$$x_1 = \sqrt{g}\, y_i\,,\quad x_i = e^{-b\Delta t} x_{i-1} + \sqrt{g\,(1-e^{-2b\Delta t})} y_i\,, (2 < i \le n)\,,$$

则 $x_1,\cdots,x_n$ 就是 OU 过程在所要的采样时刻上的值．

例１２．９　Brown 运动是时齐的 Markov 过程．其转移密度为

$$p(t,x,y) = \frac{1}{\sqrt{2pt}} e^{-\frac{(y-x)^2}{2t}}\,.$$

此时有

$$\lim_{t\to\infty} p(t,x,y) = 0\,. \tag{12．9}$$

定义１２.１０（不变密度）　若概率分布密度 $j(\cdot)$ 关于时齐 Markov 过程的转移函数 $p(t,x,y)$，满足

$$j(y) = \int p(t,x,y) j(x) dx\,,$$

则称 $j(\cdot)$ 为 $p(t,x,y)$ 的(或随机过程 $\{x_t : t \ge 0\}$ 的)不变概率分布密度函数,简称不变密度.

2　鞅列与鞅

2．1　条件期望再访(仅作参考)

背景与例子

定义１２.１１　若 (x,h) 为离散的或具有联合密度的随机向量，那么 $j(y) \overset{\Delta}{=} E(x \mid h = y)$ 是一个 Borel 函数．定义 $E(x \mid h) \overset{\Delta}{=} j(h)$．于是，对于 h 的任意的一个 Borel 函数 $h(h)$，只要它满足：对于任意 Borel 函数 g，恒有

$$E(xg(h)) = E[h(h)g(h)] \tag{E.1}$$

(这个式子等价于：对于任意 Borel 集合 Λ，恒有

$$E(xI_\Lambda(h)) = E[h(h)I_\Lambda(h)]\,), \tag{E.1'}$$

那么就有

$$P(E(x \mid h) = h(h)) = 1\,.$$

引理１２.１２（均方最佳近似）

若 $Ex^2 < \infty$，则存在 Borel 函数 $h(x)$，使

$$E(x-h(h))^2=\inf_{\text{函数}f}E(x-f(h))^2 \text{.}$$

这个 $h(h)$ 是已知 h 时 x 的最佳近似，也可以由上面的方程(E.1)或(E.1)’得到，它就是 $E(x\,|\,h)$.

定义１２.１３　（条件期望的抽象定义）

若只假定 $E\,|x\,|<\infty$，则测度论证明了，存在 Borel 函数 j 满足：
对于任意 Borel 函数 g，恒有

$$E(xg(h))=E[j\,(h)g(h)]$$

(等价于用 $\{I_\Lambda(h):\Lambda\text{为Borel集}\}$ 代替 $\{g(h):g\text{为Boral函数}\}$)．我们定义 $E(x\,|\,h)\overset{\Delta}{=}j\,(h)$．这个 j 在下述意义下唯一：若另有一个 Borel 函数 y 也满足：对于任意 Borel 函数 g 有 $E(xg(h))=E[y\,(h)g(h)]$，那么必有 $P(j\,(h)=y\,(h))=1$． 因此，条件期望 $E(x\,|\,h)$ 在不计概率为 0 的差异时是确切定义的．

显见定义１２．１１是定义１２．１３的特殊情形． 类似地可以定义 $E(x\,|\,h,V)$，$E(x\,|\,h_1,\cdots,h_n)$．

条件期望的主要性质

(1) $E(x\,|\,h)$ **对 x 具有线性性质；对常数 c 有 $E(a\,|\,h)=a$；**

(2) $E(xg(h)\,|\,h)=g(h)E(x\,|\,h)$；

(3)　若 x,h 相互独立，则 $E(f\,(x,h)\,|\,h)=[Ef\,(x,a)]_{a=h}$；

(4) $E[E(x\,|\,h,V)\,|\,h]=E(x\,|\,h)$；$E[E(x\,|\,h)\,|\,h,V]=E(x\,|\,h)$．

条件期望的定义的再推广

定义１２．１４　假定 $E\,|x\,|<\infty$，则 $E(x\,|\,h_1,\cdots,h_n,\cdots)$ 定义为：$n\to\infty$ 时 $E(x\,|\,h_1,\cdots,h_n)$ 在以下意义下的极限随机变量：

$$\lim_{n\to\infty}E\,|\,E(x\,|\,h_1,\cdots,h_n)-E(x\,|\,h_1,\cdots,h_n,\cdots)\,|=0\text{.}$$

定义１２.１５　$E(x\,|\,h_s,s\le t)$ 定义为满足下述两个条件的随机变量 $\hat{x}$：

(1) $\hat{x}\in\overline{\Phi}\,(h)$（回忆起第１１章第 2 节中的定义，若随机变量族 $h\overset{\Delta}{=}\{h_s:s\le t\}$ 中任意有限个元素的任意有界连续函数全体组成的集合为 $\Phi\,(h)$，而包含 $\Phi\,(h)$ 且对 L^2 中收敛性封闭的最小集合就是 $\overline{\Phi}\,(h)$）；

（2）对于有界的随机变量$V\in\overline{\Phi}(h)$，都有

$$E(xV)=E(\hat{x}V).$$

这种条件期望仍满足前面所述的相应性质.

*[注]　条件期望的进一步推广（对于$s-$代数F的条件期望）

1．F称为$s-$代数（或称为事件体），如果它满足：

(F.1) 必然事件$W\in F$;

(F.2) 若随机事件$A\in F$，则其对立事件$A^c\in F$;

(F.3) 若随机事件$A_n\in F$，则其并事件$\bigcup_{n=1}^{\infty}A_n\in F$.

2．若存在$\hat{x}\in F$，使得对于任意$V\in F$满足：$E(xV)=E(\hat{x}V)$，则$\hat{x}$称为x关于$s-$代数F的条件期望（在不计零概率的差异下，它也是完全确定的），我们把它记为$E(x\,|F)$．即$E(x\,|F)\in F$，且对于任意$V\in F$满足：$E(xV)=E(\,E(x\,|F)\,V)$.

3．若F是$\{h_s:s\le t\}$生成的$s-$代数，意即它是包含所有形如$\{h_s\le x\}$(任意$s\le t$，任意实数x）的事件的最小$s-$代数，则有定义可知

$$E(x\,|F)=E(x\,|h_s:s\le t).$$

这与定义１２.１５一致．　所以这里的定义是前面所有的定义的推广．

类似地有

(1)．$E(x\,|F)$对x具有线性性质，且对常数c有$E(c\,|F)=c$．

(2)．若$h\in F$（意即：对于任意实数x，事件$\{h\le x\}\in F$），则$E(xh\,|F)=hE(x\,|F)$．

(3)．若x与$s-$代数F相互独立（意即：F中的任意事件都与随机变量x独立），则$E(x\,|F)=Ex$．

(4)．若另有$s-$代数G满足：$G\subset F$（即G比F小），那么我们有

$$E[E(x\,|F)\,|G)=E(x\,|G)\,;\,E[E(x\,|G)\,|F)=E(x\,|G).$$

此外，类似地还有最佳近似性质.

2．1　鞅列

定义１２．１６　把一个随机序列$\{Y_n:n\ge 0\}$作为一系列“历史事件”的参照．记$\mathbf{Y_n}\overset{\Delta}{=}\{Y_0,\cdots,Y_n\}$，则随机变量集合$\overline{\Phi}(\mathbf{Y_n})$(即其均方信息空间)称为$n-$前历史，代表

$n-$前有用信息．另一个随机序列$\{X_n : n \ge 0\}$称为(Y_n)可知的，如果对于任意n，都有

$X_n \in \overline{\Phi}(\mathbf{Y_n})$． 随机变量$h$称为$\mathbf{Y_n}$可知的，如果$h \in \overline{\Phi}(\mathbf{Y_n})$．

注　这里的术语"可知的"，是一种直观的说法，在更为强调理论的书本或文献中，将代之以"适应的"．

定义１２．１７

随机变量序列Y_n 称为鞅列，如果对$\forall n, E|Y_n| < \infty$，$\forall m$有

$$E(Y_{n+m} \mid \mathbf{Y_n}) = Y_n; \qquad (12.10)$$

又X_n称为 (Y_n)鞅列，如果$\{X_n\}$为(Y_n)可知的，且对$\forall n, E|X_n| < \infty$，$\forall m$有

$$E(X_{n+m} \mid \mathbf{Y_n}) = X_n. \qquad (12.11)$$

由于$\{X_n\}$为(Y_n)可知的，由条件期望的性质得到，若X_n是(Y_n)鞅列，则X_n也是鞅列．

［注］ 如果上面式子中的 "=" 分别改为 "≥" 或 "≤"，则对应的随机序列分别称为下鞅列，或上鞅列．

命题１２．１８ $\forall n, m, E(X_{n+m} \mid \mathbf{Y_n}) = X_n$ ，当且仅当． $\forall n, E(X_{n+1} \mid \mathbf{Y_n}) = X_n$．

即X_n为(Y_n)鞅，等价于，对于任意n有

$$E(X_{n+1} \mid \mathbf{Y_n}) = X_n. \qquad (12.11)'$$

证明　对m用归纳法，由条件期望的性质立得

$$E(X_{n+m+1} \mid \mathbf{Y_n}) = E(\,E(X_{n+m+1} \mid \mathbf{Y_{n+1}}) \mid \mathbf{Y_n}) = E(X_{n+1} \mid \mathbf{Y_n}) = X_n.$$

鞅列的直观含义．

假定Y_n是鞅列．设想我们在时刻n以本金Y_n参加博弈，在下一时刻$n+1$的实际资金Y_{n+1}是随机的，它在"最近"的历史Y_n已知条件下的条件期望$E(Y_{n+1} \mid Y_n)$为

$$E(Y_{n+1} \mid Y_n) = E(E(Y_{n+1} \mid \mathbf{Y_n}) \mid Y_n) = E(Y_n \mid Y_n) = Y_n.$$

即在Y_n已知的条件下，下一时刻$n+1$的收益的条件平均与现有的本金Y_n恰好相等．这说明了此博弈是(平均)公平的．因此鞅列代表一个平均趋势公平发展的博弈的本金随机序列．而下鞅列则代表按平均趋势必赢的博弈，上鞅列则代表按平均趋势必输的博弈．

例１２．１９（随机徘徊） $Y_0 = 0, Y_n = x_1 + \cdots + x_n, \{x_i\}$独立同分布，且$Ex_i = 0$．$Y_n$即随机徘徊．易验证它是鞅列．

例１２．２０． $X_n \overset{\Delta}{=} E(X \mid \mathbf{Y_n})$是$(Y_n)$鞅列．

例１２．２１ 如果$\{X_n\}$为(Y_n)可知的，而 $Z_n = X_n - \sum_{k=0}^{n-1}[E(X_{k+1} \mid \mathbf{Y_k}) - X_k]$，则易检查$(Z_n\}$为$(Y_n)$鞅列． Z_n正是由X_n相继地减去它与鞅列的差的波动而得到的，因此在直观上它应该是鞅列．(事实上我们有

$$E(Z_{n+1} \mid \mathbf{Y_n}) = E(X_{n+1} \mid \mathbf{Y_n}) - \sum_{k=0}^{n}(E[E(X_{k+1} \mid \mathbf{Y_k}) \mid \mathbf{Y_n}] - E(X_k \mid \mathbf{Y_n}))$$

$$= E(X_{n+1} \mid \mathbf{Y_n}) - \sum_{k=0}^{n}(E(X_{k+1} \mid \mathbf{Y_k}) - X_k)$$

$$= X_n - \sum_{k=0}^{n-1}(E(X_{k+1} \mid \mathbf{Y_k}) - X_k) = Z_n.$$

例１２．２２ (平方可积鞅列). 在例 12.19 中若$Ex_i^2 = s^2 < \infty$，定义

$$Z_n = Y_n^2 - EY_n^2 = Y_n^2 - ns^2,$$

则Z_n为(Y_n)鞅列.

证明 令$X_n = Y_n^2$，则X_n为(x_n)可知的. 记$\mathrm{X_n} = \{x_0, \cdots, x_n\}$. 因为$x_{k+1}$与$\mathrm{X_k}$独立, 所以$E(x_{k+1} \mid \mathrm{X_k}) = Ex_{k+1} = 0$，$E(x_{k+1}^2 \mid \mathrm{X_k}) = Ex_{k+1}^2 = s^2$. 于是

$$E(X_{k+1} \mid \mathrm{X_k}) - X_k = E(Y_{k+1}^2 \mid \mathrm{X_k}) - Y_k^2 = E((Y_k + x_{k+1})^2 \mid \mathrm{X_k}) - Y_k^2$$

$$= [Y_n^2 + 2E((Y_k x_{k+1}) \mid \mathrm{X_k}) + s^2] - Y_k^2 = 2Y_k E(x_{k+1} \mid \mathrm{X_k}) + s^2 = s^2.$$

可见有 $E(Z_{n+1} \mid \mathrm{X_n}) = E(X_{n+1} \mid \mathrm{X_n}) - (n+1)s^2 = X_n^2 - ns^2 = Z_n$，所以$Z_n$为$(x_n)$鞅.

例１２．２３ (随机利率) 设X_n为(Y_n)可知的非负随机序列. 对$\mathbf{Y_n} = \{Y_0, \cdots, Y_n\}$，若随机利率$d_n$是$(Y_n)$可知的, 且满足条件

$$E(X_{n+1} \mid \mathbf{Y_n}) = e^{d_n} X_n.$$

令

$$Z_0 = 1,\ Z_n = e^{-(d_0 + \cdots + d_{n-1})} X_n.$$

那么Z_n为(Y_n)鞅列.

证明 $E(Z_{n+1} \mid \mathbf{Y_n}) = e^{-(d_0 + \cdots + d_n)} E(X_{n+1} \mid \mathbf{Y_n}) = Z_n$.

例１２．２４ （**指数鞅列**） **在例１２．１９中令** $Z_n = \prod_{k=1}^{n} \frac{e^{sx_k}}{Ee^{sx_k}}$，**则** Z_n **为** (x_n) **鞅列**.

证明　**记** $\mathrm{X}_{\mathbf{n}} = \{x_0, \cdots, x_n\}$．**由** x_{n+1} **与** $\mathrm{X}_{\mathbf{n}}$ **的独立性立得**

$$E(Z_{n+1} \mid \mathrm{X}_{\mathbf{n}}) = Z_n \frac{1}{Ee^{sx_{n+1}}} E(e^{sx_{n+1}} \mid \mathrm{X}_{\mathbf{n}}) = Z_n .$$

例１２．２５（分支鞅列）　**设开始时刻第 0 代有一个细胞，在下一个整数时刻分裂为第一代的细胞后，各个新细胞又按相同的规律独立地在下一个整数时刻再进行同样方式的分裂．记第** n **代第** k **个细胞分裂的个数为随机变量** $h_{n,k}$，**它们是独立同分布的，且假定它们有期望** m．**设经过第** n **代分裂后的细胞总数为** x_n，**于是**

$$x_0 = 1, x_n = h_{n,1} + \cdots + h_{n,x_{n-1}} .$$

那么 $\frac{x_n}{m^n}$ **是鞅列．**

证明　**由各代之间以及同一代的各个个体间的独立性，我们有**

$$E\left(\frac{x_n}{m^n} \mid x_0, \cdots, x_{n-1}\right) = E\left(\frac{h_{n,1} + \cdots + h_{n,x_{n-1}}}{m^n} \mid x_0, \cdots, x_{n-1}\right)$$

$$= E\left(\frac{h_{n,1} + \cdots + h_{n,x_{n-1}}}{m^n} \mid x_{n-1}\right) = \frac{1}{m^n} \sum_{k=1}^{x_{n-1}} Eh_{n,k} = \frac{x_{n-1} m}{m^n} = \frac{x_{n-1}}{m^{n-1}} .$$

可见结论成立．

例１２．２６（似然比）　**设** $\{Y_n\}$ **为独立同分布列，具有分布密度** g．**若** $f(x) \ge 0$ **满足** $\int f(x)dx = 1$（**即** $f(x)$ **是另一个随机变量的分布密度**），**那么**

$$X_n \overset{\Delta}{=} \prod_{k=1}^{n} \frac{f(Y_k)}{g(Y_k)}$$

是 (Y_n) **鞅列．**

证明　**因为** $E\left(\frac{f(Y_n)}{g(Y_n)}\right) = \int \frac{f(y)}{g(y)} g(y)dy = 1$，**所以有**

$$E(X_n \mid Y_1, \cdots, Y_{n-1}) = X_{n-1} E\left(\frac{f(Y_n)}{g(Y_n)} \mid Y_1, \cdots Y_{n-1}\right) = X_{n-1} E\left(\frac{f(Y_n)}{g(Y_n)}\right) = X_{n-1} .$$

例１２．２７（Levy 鞅）

设时齐的 Markov 链 x_n **的转移矩阵为** $\mathbf{P} = (p_{ij})$．**对于状态空间上的任意有界函数** f，

定义有界函数 f 到有界函数的映射 $\mathbf{P} f$ 为

$$(\mathbf{P} f)(i)=\sum_j p_{ij} f(j).$$

对于给定的函数 $f(i)$，令

$$X_n = f(x_n)-f(x_0)-\sum_{k=0}^{n-1}(\mathbf{P}-\mathbf{I})f(x_k),$$

其中 $\mathbf{I}$ 为恒同映射，即对于任意 f，都有 $\mathbf{I} f = f$．那么 X_n 是 (x_n) 鞅．

证略.

鞅列的性质，停时与选择定理

命题１２.２８　设 X_n 为 (Y_n) 鞅列，则 $EX_n=$ 常数．

证明　用全期望定理 $EX_{n+1}=E[E(X_{n+1}\mid \mathbf{Y_n})]=EX_n$.

定义１２.２９　一个可以取值 $+\infty$ 的非负整值随机时刻 t，称为关于 (Y_n) 的停时（简称 (Y_n) 停时)，如果事件 t 在 n 时发生与否，只依赖于 (Y_n) 的历史记录 $\mathbf{Y_n}=\{Y_0,\cdots,Y_n\}$，即复合随机变量 $I_{\{n\}}(t)$ 是 $\mathbf{Y_n}$ 可知的随机变量．

在博弈过程中，参与者常根据他事先设计的规则，来决定退出与否(典型的规则，例如有:，输完退出，输赢超过本钱的 80%时退出等等．这些事先设计好的时刻都是随机的，根据这些设计好的退出时刻,可知"在时刻 n 退出" 是一个随机事件，它由时间序列在时刻 n 与时刻 n 以前的历史情况所完全确定．因此，这些设计好的随机时刻都是停时)，这种退出时刻就是停时.

命题１２.３０　固定的非负整值常数 n 是一个 (Y_n) 停时；又如果 t,s 都是 (Y_n) 停时，$n\ge 0$，则 $t\wedge s$，$t\wedge n$ 也是 (Y_n) 停时.

在此，我们要解释清楚一个记号．随机序列 $\{X_n)$ 之所以可以有有限维分布的概念，是因为这时作一次试验所得的基本事件 w 确定后，随机序列 $\{X_n)$ 就相应地得到了采样数值列 $\{X_n(w)\}$．假定 t 是一个只取有限值的停时，那么 $t=t(w)$ 就是作一次试验所得的基本事件 w 确定后一个时刻的值．此时，我们可以自然的定义随机序列 $\{X_n\}$ 在停时 t 上的取值为：$X_t \overset{D}{=} X_{t(w)}(w)$.

下面的采用的选样定理的叙述形式，是概率论中非常著名的 Doob 停止定理的简单部分，

它在应用中是常见的(Doob 停止定理的一般形式则需要较多的工具，本书不再列入)．

如上所述，典型的停时是参加博弈的人给自己预先制定的退出博弈的时间 t ．例如，他的资金达到 1000 元的时刻(若在时刻 n，他有 x_n 元，那么，t 就是 x_n 首达 1000 的时刻)．一个鞅所描述的博弈是公平的，即对于任意 n 均有 $Ex_n = Ex_0$．那么在任意停时 t 上是否也有 $Ex_t = Ex_0$ 呢？一般这并不一定成立．例如，x_n 是满足 $x_0 = 1$ 的随机徘徊，令停时 t 为首达 2 的时刻．由于简单随机徘徊 x_n 可以取负值，这就是容许参加博弈的人欠钱，但是由于每次输赢的概率都是 $\frac{1}{2}$，这时有 $P(t<\infty)=1$．于是我们有 $Ex_0 = 1, Ex_t = 2$．即博弈人在时刻 t 停止的策略是稳赢 1 元的策略，因此它不是公平策略，其原因在于：让参加博弈的人不加限制地欠钱是一个对于他过分有利的条件，这对于他的博弈对手是很不公平的．由此看来要满足 $Ex_t = Ex_0$，停时 t 就必须有所限制．下面的定理给出了常见的充分条件．

定理１２.３１　(选样定理)

设 X_n $(n\geq 0)$ 是鞅．若 t 是一个（X_n）停时，且满足下面两个条件之一：

（1）t 是有界停时，

（2）$E|X_t|<\infty$ 且 $\lim_{n\to\infty} E(X_n\cdot I_{\{t>n\}})=0$，

那么我们有

$$EX_t = EX_0 .$$

证明　先假定 t 有界．设 $t\leq m$．由于 t 是停时，所以随机事件 $\{t\geq k\}=\{t\leq k-1\}^c$ 是由 $X_{k-1},\cdots,X_0$ 的信息确定的，于是有

$$E[(X_k - X_{k-1})I_{t\geq k}\,|\,X_{k-1},\cdots,X_0] = I_{t\geq k}E[(X_k - X_{k-1})\,|\,X_{k-1},\cdots,X_0]=0 .$$

由此我们得到

$$E(X_t - X_0)=\sum_{j=1}^{m} E[(X_j - X_0)I_{t=j}] = \sum_{j=1}^{m} E[\sum_{k=1}^{j} (X_k - X_{k-1})I_{t=j}]$$

$$=\sum_{k=1}^{m} E[(X_k - X_{k-1})\sum_{j=k}^{m} I_{t=j}] = \sum_{k=1}^{m} E[(X_k - X_{k-1})I_{t\geq k}]=0 .$$

对于一般的 t．那么 $t_m = t\wedge m$ 也是停时，故 $EX_{t_m} = EX_0$．从而由假定推出

$$|EX_t - EX_{t_m}| = |E[(X_t - X_m)I_{t\geq m}]\leq |E(X_t I_{t\geq m})|+|E(X_m I_{t\geq m}| \xrightarrow{m\to\infty} 0 .$$

最后就得到 $EX_t = EX_0$．

[注]　条件$\lim_{n\to\infty} E(x_n \cdot I_{\{t>n\}})=0$是一个纯粹技术性的条件.

选择定理告诉我们: 对于一个(平均地)公平的博弈, 它在＂公平的停时＂上也反映了(平均地)公平的性质.

鞅方法与选择定理在应用中的典型例子是博弈输光问题.

例１２.３２　(输光问题)　设甲, 乙两人作双人博弈, 每次 1 元, 直至其中一人输光为止. 设在时刻 0 甲有初始资金a元,乙有初始资金b元. 假定每次博奕时甲赢的概率是p，而乙赢的概率是$q=1-p$. 求甲先输光的概率p_a, 及乙先输光的概率q_b，并求此种博奕的平均持续时间.

记时刻n时甲有资金x_n元. 输光问题实际上是一个简单随机徘徊被两个吸收壁０与$a+b$中那一个吸收的问题（甲输光对应于被０吸收，乙输光对应于被$a+b$吸收）. 令x_n为简单随机徘徊：$x_n=a+X_1+\cdots+X_n, \{X_n: n\geq 0\}$独立同分布, 且$X_n \sim \begin{pmatrix} -1 & 1 \\ q & p \end{pmatrix}$, 令$T=\inf\{n: x_n=0\text{或}a+b\}$($\inf\quad f\quad=+\infty$). 于是

$$p_a=P(x_T=0), q_b=P(x_T=a+b), p_a+q_b=1.$$

为了利用鞅方法求p_a, 我们区分如下各种情形:

(1)　$p=q$ 情形.

我们有$E(x_{n+1}\mid x_n,\cdots, x_0)=p(x_n+1)+q(x_n-1)=x_n$, 所以$x_n$是鞅. 注意到有双侧吸收壁$0, a+b$的简单随机徘徊作为 Markov 链时只有$0, a+b$是常返态，故$T<\infty$. 于是$E(|x_n| I_{\{T>n\}})\leq (a+b)EI_{\{T>n\}}=(a+b)P(T>n)\to 0$. 于是由定理１２．３１得到$Ex_T=Ex_0=a$. 但是

$$Ex_T=0\cdot P(x_T=0)+(a+b)P(x_T=a+b)=(a+b)q_b.$$

从而$q_b=\dfrac{a}{a+b}$，这是乙输光的概率. 进而有$p_a=1-q_b=\dfrac{b}{a+b}$.

下面求此博奕的平均持续时间$E(T\mid x_0=a)$. 为此我们定义$h_n=x_n^2-n$. 因为

$$E(h_{n+1}\mid x_n,\cdots,x_0)=E(x_{n+1}^2\mid x_n,\cdots,x_0)-(n+1)$$

$$=p(x_n+1)^2+q(x_n-1)^2-(n+1)=x_n^2-n=h_n,$$

所以h_n是(x_n)鞅. 利用直观, 我们指出$ET<\infty$ (事实上, 假定V_n是以 0 与$a+b$为反射壁的反射

随机徘徊，那么V_n是正常返的.再记$T'=\inf\{n: V_n=0$或$a+b\}$，就有$ET'<\infty$.由此得到$ET<ET'<\infty$). 由$|h_n|\le (a+b)^2+n$，$T<\infty$及$E(nI_{T>n})<E(TI_{T>n})\to 0(n\to\infty)$推出$\lim_{n\to\infty}E(h_n\cdot I_{\{T>n\}})=0$. 用定理１２．３１得到$Eh_T=Eh_0=a^2$. 即

$$a^2=Ex_T^2-ET=0\cdot p_a+(a+b)^2q_b-ET=(a+b)a-ET.$$

由此得到

$$ET=ab.$$

(2)　$p>q$情形，令$m=p-q$.

这时是不公平的博弈,甲方赢的概率大. 为了计算p_a,q_b，令$V_n=(\frac{q}{p})^{x_n}$. 此时$0<\frac{q}{p}<1$.

我们验证V_n是(x_n)鞅. 事实上

$$E(V_{n+1}|x_n,\cdots,x_0)=E[(\frac{q}{p})^{x_{n+1}}|x_n,\cdots,x_0]=p(\frac{q}{p})^{(x_n+1)}+q(\frac{q}{p})^{(x_n-1)}=(\frac{q}{p})^{x_n}=V_n.$$

类似地,我们也有$\lim_{n\to\infty}E(V_n\cdot I_{\{T>n\}})=0$.同样地由定理１２．３１推出$EV_T=EV_0=(\frac{q}{p})^a$.

另一方面按定义我们还有

$$EV_T=(\frac{q}{p})^0p_a+(\frac{q}{p})^{a+b}(1-p_a).$$

把它代入前一式，我们得到

$$p_a=\frac{(\frac{q}{p})^a-(\frac{q}{p})^{a+b}}{1-(\frac{q}{p})^{a+b}},\quad q_b=\frac{1-(\frac{q}{p})^a}{1-(\frac{q}{p})^{a+b}}.$$

类似地，如果$x_0=0$，　$t[-a,b]=\inf\{x_n=-a$或$b\}$，p_a,q_b定义如前. 那么

$$P(x_{t[-a,b]}=-a)=p_a,\ P(x_{t[-a,b]}=b)=q_b.$$

为了求此博奕的平均持续时间，我们再定义$Z_n=x_n-nm$. 它满足

$$E(Z_{n+1}|x_n,\cdots,x_0)=E(x_{n+1}|x_n,\cdots,x_0)-(n+1)m$$

$$=p(x_n+1)+q(x_n-1)-(n+1)m=Z_n,$$

即Z_n是(x_n)鞅. 类似地可以验证$\lim_{n\to\infty}E(Z_n\cdot I_{\{T>n\}})=0$. 我们由定理１２．３１推出

$EZ_T = EZ_0$，即 $0 \cdot p_a + (a+b)q_b - mET = a$．故而有

$$ET = \frac{1}{m}[(a+b)q_b - a]\text{.}$$

下面我们假定$x_0 = 0$．记x_n首达 1 的时刻为t_1．我们要计算Et_1．为此令

$$t[-N,1] = \inf\{ x_n = -N \text{或} 1\}\text{，}$$

它与t_1都是(x_n)停时．类似地验证$\lim_{n\to\infty} E(Z_n \cdot I_{\{t[-N,1]>n\}}) = 0$．于是由定理１２．３１得

$EZ_{t[-N,1]} = EZ_0$，即

$$-Np_N + 1\cdot(1-p_N) - mEt[-N,1] = 0\,,$$

此处 $p_N = P(Z_{t[-N,1]} = -N)$．可见

$$Et[-N,1] = \frac{1}{m}[1-(N+1)p_N] = \frac{1}{m}[1-(N+1)\frac{(\frac{q}{p})^N - (\frac{q}{p})^{N+1}}{1-(\frac{q}{p})^{N+1}}].$$

于是我们有

$$Et_1 = \lim_{N\to\infty} Et[-N,1]\frac{1}{m} = \frac{1}{p-q}\text{.}$$

(3) $p<q$ 情形，留作习题.

定理１２.３３（下鞅列的 Doob 分解）

设X_n是(Y_n)下鞅列，则存在唯一的(Y_n)鞅列M_n，满足:

（1）初值$M_1 = X_1$.

（2）存在(Y_{n-1})可知的（也称为(Y_n)可料的）递增随机序列A_n，使$A_1 = 0$，且

$$X_n = M_n + A_n. \qquad (11.\ 14)$$

证明　存在性: 令

$$A_n = \sum_{k=1}^{n-1}[E(X_{k+1} \mid \mathbf{Y_k}) - X_k], (n\ge 2), A_1 = 0\,.$$

由此定义立见随机序列 A_n是(Y_{n-1})可知的.　用下鞅性得到 A_n是递增的. 再令

$$M_n = X_n - \sum_{k=1}^{n-1}[E(X_{k+1} \mid \mathbf{Y_k}) - X_k]，(n\ge 2)，M_1 = X_1,$$

则容易验证它是鞅列.

唯一性: 设另有一个符合条件的分解 $X_n = M_n' + A_n'$. 那么鞅列

$$M_n - M_n' = A_n' - A_n$$

应该是 (Y_{n-1}) 可知的. 所以

$$M_n - M_n' = E[(M_n - M_n') \mid \mathbf{Y_{n-1}}) = M_{n-1} - M_{n-1}' = \cdots = M_1 - M_1' = 0.$$

2. 2 **连续时间参数的鞅**

定义１２．３４

用一个随机过程 $\{Y_t : t \geq 0\}$ 作为整个"历史"的参照．另一个随机过程 $\{X_t : t \geq 0\}$ 称为 (Y_t) 可知的，如果对于任意 t, X_t 可以表示成 $\{Y_s : s \leq t\}$ 中元素的函数的极限．

随机过程 Y_t 称为鞅，如果任意 t 有 $E|Y_t| < \infty$，且对于任意 n，任意 $t > s > s_1 > \cdots > s_n$ 有

$$E(Y_t \mid Y_s, Y_{s_1}, \cdots, Y_{s_n}) = Y_s. \qquad (12.15)$$

X_t 称为 (Y_t) 鞅，如果 $\{X_t\}$ 为 (Y_t) 可知的，$E|X_t| < \infty$，且对于任意 n，任意 $t > s > s_1 > \cdots > s_n$ 有

$$E(X_t \mid Y_s, Y_{s_1}, \cdots, Y_{s_n}) = X_s. \qquad (12.16)$$

由条件期望的性质，可以证明这个定义等价于：对于任意 $t > s$ 有

$$E(X_t \mid Y_u : u \leq s) = X_s. \qquad (12.16)'$$

由于 $\{X_t\}$ 为 (Y_t) 可知的，由条件期望的性质得到，若 X_t 是 (Y_t) 鞅，则 X_t 也是鞅．

与鞅列类似，连续时间的鞅的直观含义仍然是公平博弈．

例１２．３５　独立增量过程 Y_t，如果满足 $Y_0 = 0, \forall t, EY_t = 0$，那么它是鞅．

证明　注意 $Y_0 = 0$，故与 $(Y_s - Y_{s_1}, Y_{s_1} - Y_{s_2}, \cdots, Y_{s_n} - Y_0)$ 独立，而后者又与 $(Y_s, Y_{s_1}, \cdots, Y_{s_n})$ 相互线性表出，所以 $Y_t - Y_s$ 就与 $(Y_s, Y_{s_1}, \cdots, Y_{s_n})$ 独立. 用条件期望的性质，对于任意 n，任意 $t > s > s_1 > \cdots > s_n$，有

$$E(Y_t \mid Y_s, Y_{s_1}, \cdots, Y_{s_n}) = E(Y_t - Y_s \mid Y_s, Y_{s_1}, \cdots, Y_{s_n}) + E(Y_s \mid Y_s, Y_{s_1}, \cdots, Y_{s_n})$$

$$= E(Y_t - Y_s) + Y_s = Y_s.$$

推论１２．３６　Brown 运动B_t是鞅；若N_t为强度为 l 的 Poisson 过程，则 $N_t - lt$ 是鞅.

例１２．３７（平方可积鞅）　设例１２．３５中的独立增量过程有二阶矩，即对于任意 t，令$EY_t^2 \overset{D}{=} s_t^2 < \infty$，$Z_t \overset{\Delta}{=} Y_t^2 - s_t^2$，则$Z_t$为$(Y_t)$鞅.

证明　显然 X_t 为 (Y_t) 可知的．由例１２．３５中的分析，$(Y_t - Y_s)^2$ 也与 $(Y_s, Y_{s_1}, \cdots, Y_{s_n})$ 独立．因而，对于任意n，任意$t > s > s_1 > \cdots > s_n$，有

$$E(Y_t - Y_s \mid Y_s, Y_{s_1}, \cdots, Y_{s_n}) = E(Y_t - Y_s) = 0,$$

于是

$$E(Y_t^2 - Y_s^2 \mid Y_s, Y_{s_1}, \cdots, Y_{s_n}) = E([(Y_t - Y_s)^2 + 2(Y_t - Y_s)Y_s] \mid Y_s, Y_{s_1}, \cdots, Y_{s_n})$$

$$= E(Y_t - Y_s)^2 + 2Y_s E(Y_t - Y_s) = E(Y_t - Y_s)^2$$

$$= EY_t^2 - EY_s^2 - 2E[Y_s(Y_t - Y_s)] = s_t^2 - s_s^2 .$$　　　　】

推论１２．３８　对于 Brown 运动B_t而言，$B_t^2 - t$是(B_t)鞅；又若N_t为强度为 l 的 Poisson 过程，则 $(N_t - lt)^2 - lt$是(N_t)鞅.

例１２．３９　（指数鞅）　对 Brown 运动B_t，$z_t = e^{cB_t - \frac{c^2 t}{2}}$是$(B_t)$鞅.

证明　任意n，任意$t > s > s_1 > \cdots > s_n$，由例１２．３５知道$\frac{z_t}{z_s} = e^{c(B_t - B_s) - \frac{c^2(t-s)}{2}}$与 $(B_s, B_{s_1}, \cdots, B_{s_n})$独立．再注意$z_s$是 B_s 的函数，利用条件期望的性质便得

$$E(z_t \mid B_s, B_{s_1}, \cdots, B_{s_n}) = z_s E(\frac{z_t}{z_s} \mid B_s, B_{s_1}, \cdots, B_{s_n})$$

$$= z_s E(e^{c(B_t - B_s) - \frac{c^2(t-s)}{2}} \mid B_s, B_{s_1}, \cdots, B_{s_n}) = z_s E(e^{c(B_t - B_s) - \frac{c^2(t-s)}{2}}) = z_s .$$

例１２．４０　（与连续时间的 Markov 链相系的鞅）

设$\{x_t, t \geq 0\}$为具有保守的转移速率矩阵$Q = (q_{ij})$的连续时间的 Markov 链．假定它具有右连续的轨道(即w 固定式时，$x_t(w)$是t的右连续函数)．那么如下定义的随机过程

$$X_t = I_{\{j\}}(x_t) - \int_0^t q_{x_u j} du , \qquad (12.17)$$

$$Y_t = I_\Lambda(x_t) - \int_0^t q_{x_u\Lambda} du , \tag{12. 18}$$

$$Z_t = f(x_t) - \int_0^t (Qf)(x_u) du \tag{12. 19}$$

都是$x-$鞅, 其中Λ是一个状态集， $q_{i\Lambda} = \sum_{j\in\Lambda} q_{ij}$，函数$f(i)$只在有限个$i$上取非 0 值, 而

$$(Qf)(i) = \sum_j q_{ij} f(j) .$$

(证明提示　对于 任意$t > s$,利用 Markov 性，我们有(把s看成现在，则$\int_s^t q_{x_u j} du$是过程的将来的值)

$$E(X_t | x_u : u \le s) = E(I_{\{j\}}(x_t | x_s) - E(\int_0^s q_{x_u j} du | x_u : u \le s) - E(\int_s^t q_{x_u j} du | x_s)$$

$$= p_{x_s j}(t-s) - \int_0^s q_{x_u j} du - \int_s^t E(q_{x_u j} | x_s) du$$

$$= p_{x_s j}(t-s) - \int_0^s q_{x_u j} du - \int_s^t \sum_k q_{kj} P(x_u = k | x_s) du$$

$$= p_{x_s j}(t-s) - \int_0^s q_{x_u j} du - \int_s^t \sum_k p_{x_s k}(u-s) q_{kj} du$$

$$= p_{x_s j}(t-s) - \int_0^s q_{x_u j} du - \int_s^t p_{x_s j}'(u-s) du \quad (\text{Master 方程})$$

$$= p_{x_s j}(0) - \int_0^s q_{x_u j} du = I_{\{j\}}(x_s) - \int_0^s q_{x_u j} du = X_s .$$

类似地可以证明Y_t, Z_t也都是鞅.

命题１２.４１　设X_t为 (Y_t)鞅，则 $EX_t \equiv EX_0$.

定义１２.４２　可取值$+\infty$的非负随机时刻t ，称为关于(Y_t)的停时 (简称(Y_t)停时), 如果事件t 在t前发生与否只依赖于(Y_t)的历史记录, 即随机过程$I_{\{t\le t\}}$是(Y_t)可知的. 随机变量h称为$\{Y_s : s \le t\}$可知的，如果h可以表示成$\{Y_s : s \le t\}$中的元素的函数列的极限.

注　同样,”可知的”是直观的术语,在学术名称为”适应的”.

显见，常数时间是停时；又如果t, s 都是停时，$t \geq 0$，则$t \wedge s, t \wedge t$ 也是停时.

与离散时间的鞅列类似地，对连续时间的鞅同样有如下的选样定理：

定理１２．４３　（选样定理）

设$x_t\ (t \geq 0)$是轨道连续的（其实只需轨道右连续）鞅．若停时t 满足下述条件之一：

（１）t 是有界停时，

（２）$E|x_t| < \infty$ 且 $\lim_{t \to \infty} E(x_t \cdot I_{\{t > t\}}) = 0$ ，

那么我们有

$$Ex_t = Ex_0 .$$

（证略）

选样定理对研究漂移 Brown 运动是一个重要的工具．

龚光鲁，钱敏平著　应用随机过程教程 – 与在算法和智能计算中的应用
清华大学出版社，2003

下面我们粗略地介绍随机分析的基本工具，即随机微积分．还要介绍一种最常见的连续状态连续时间的 Markov 过程，称为扩散过程．由于这部分内容涉及较多的数学内容，要真正表达清楚，测度论与概率基础等工具是必不可少的．这样就大大超出了本书大部分读者现有的数学基础．因此，在本章中我们并不求给出一般的定义与定理的精确叙述，而是只从一些有典型意义的特例，来导出随机微积分及扩散过程的概念与其思想之精髓．

3．Ito 积分　－ 对 Brown 运动的积分

对于函数的研究，函数的微积分是其精髓．而对于一个随机过程的函数这样的特殊随机函数而言，其微积分也具有同样的重要性． 函数的微分与积分是一对互逆的运算． 对于初等函数，我们常从其导数入手，进而得到微分与积分；但是对非常复杂的一般函数，积分却比导数更容易理解与处理，也易于作近似计算． 因此，人们也常以积分作为微积分的核心．在本节中，我们将考虑对 Brown 运动的随机积分，作为随机微积分的核心．

3．1　对 Brown 运动的积分与其特殊性

对 Brown 运动的积分的特殊性

设在概率空间(Ω, F, P)上有 Brown 运动$\{B_t\}$（$B_t = B_t(w), w\in\Omega$）及另一个轨道(样本函数)连续的随机过程 $\Phi_t \overset{\Delta}{=} \Phi_t(w)$ （回忆起它是依赖于参数t的随机变量族，在w固定时，$\Phi_t(w)$作为t的函数，即为随机过程Φ_t的一个样本函数，或轨道)．我们能不能对于固定的w，定义样本函数$F_t(w)$对于 Brown 运动的样本函数$B_t(w)$的积分$\int_a^b \Phi_t(w)dB_t(w)$为积分和的极限呢 ？ 更具体地，如果我们考虑区间$[a,b]$的一组划分：

$$a = t_0^{(n)} < \cdots < t_{l_n}^{(n)} = b\text{，}\Delta_n \overset{\Delta}{=} \max_{0\le k\le l_n-1}(t_{k+1}^{(n)} - t_k^{(n)}) \to 0\text{，}$$

令

$$J_n(w) = \sum_{k=0}^{l_n-1} \Phi_{t_k^{(n)}}(w)(B_{t_{k+1}^{(n)}}(w) - B_{t_k^{(n)}}(w))\text{，} \qquad (12．24)$$

那么能否定义$\int_a^b \Phi_t(w)dB_t(w)$为$\lim_{n\to\infty} J_n(w)$呢 ？　如果$\lim_{n\to\infty} J_n(w)$对每一个ω都存在，那么很自然地，这个极限就该是$\Phi_t$对$B_t$的积分，但是不幸的是，一般$\lim_{n\to\infty} J_n(w)$并不存在，关于这一点，我们将在下面说明．

另一方面，如果我们不要求$\lim_{n\to\infty} J_n(w)$对每一个ω都存在，而把

$J_n \overset{\Delta}{=} J_n(w), w \in \Omega$ 视为随机变量，而且把 ″极限″ 的含义要求得弱一些：如果随机过程 Φ_t 只依赖 Brown 运动的过去 $\{B_u : u \le t\}$，即 Φ_t 为 (B_t) 可知的，则可以证明随机列 J_n 是按概率收敛的．Ito 就把这个概率收敛的极限随机变量，定义为随机过程 Φ_t 对 Brown 运动 B_t 的积分．这就是 Ito 积分的基本思想．即若对于任意 $e > 0$ 有

$$\lim_{n\to\infty} P(|J_n(w) - h(w)| > e) = 0，$$

则我们就定义

$$\int_a^b \Phi_t(w) dB_t(w) \overset{\Delta}{=} \eta(\omega).$$

并把它简记为 $\int_a^b \Phi(t) dB_t = h$，或更简单地记为 $\int_a^b \Phi dB$．

需要指出：和数 J_n 在每个小区间 $(t_k^{(n)}, t_k^{(n)}]$ 上，我们需要限定取 $\Phi_{t_k^{(n)}}$ 作为 Φ 在此小区间上的近似，而不能象在普通函数的积分中那样，可以取 Φ 在此小区间上任意的一点的值为近似．其原因是：这里相应于普通积分中的差分的项是：$\Delta B_{t_k^{(n)}} = B_{t_{k+1}^{(n)}} - B_{t_k^{(n)}}$，它们是随机变量．虽然 $\lim_{n\to\infty} \Delta B_k^{(n)}(w) = 0$，但是对于不同的 ω，它们趋于 0 的速度很不一致，而粗略地说，平均地有 $(\Delta B_k^{(n)})^2 \sim E(\Delta B_k^{(n)})^2 = \Delta t_k^{(n)}$．也就是说，平均地 $\Delta B_k^{(n)}$ 趋于 0 的速度为 $\sqrt{Dt_k^{(n)}}$，它大大地慢于 $\Delta t_k^{(n)}$．从而在区间 $(t_k^{(n)}, t_k^{(n)}]$ 上取不同的点作为 Φ 的近似，所得的近似和之间的差别是不可忽略的．

定义（命题）１２．４４　（Ito(随机)积分的定义）

若随机过程 Φ_t 是 (B_t) 可知的，且

$$\int_0^T E|\Phi_t(w)|^2 \, dt < \infty，\qquad (12.25)$$

则对于区间 $[0,T]$ 的任意一组划分：

$$0 = t_0^{(n)} < \cdots < t_{l_n}^{(n)} = T，\Delta_n \overset{\Delta}{=} \max_{0\le k\le l_n - 1}(t_{k+1}^{(n)} - t_k^{(n)}) \to 0，$$

和数

$$J_n = \sum \Phi_{t_k^{(n)}} (B_{t_{k+1}^{(n)}} - B_{t_k^{(n)}})$$

必然按概率收敛到某个随机变量，则这个极限随机变量就定义为 $\int_0^T \Phi(t) dB_t$，称为 Ito 积

分.

例12.45　我们有

$$\int_0^t B_s dB_s = \frac{1}{2}B_t^2 - \frac{1}{2}t .$$

这个结果与普通微积分的不定积分公式是不同处，是这里多了$\frac{1}{2}t$．下面我们来推导这一结论．记$\Delta B_{t_k^{(n)}} = B_{t_{k+1}^{(n)}} - B_{t_k^{(n)}}$，则

$$J_n = \sum_k B_{t_k^{(n)}} (B_{t_{k+1}^{(n)}} - B_{t_k^{(n)}})$$

$$= \frac{1}{2}\sum_k [2B_{t_k^{(n)}} B_{t_{k+1}^{(n)}} - 2B_{t_k^{(n)}}{}^2] = \frac{1}{2}\sum_k [\Delta(B_{t_k^{(n)}})^2 - (\Delta B_{t_k^{(n)}})^2]$$

$$= \frac{1}{2}B_t^2 - \frac{1}{2}\sum_k (\Delta B_{t_k^{(n)}})^2 \overset{\Delta}{=} \frac{1}{2}B_t^2 - I_n .$$

而由 Brown 运动的性质，我们有

$$EI_n = \frac{t}{2}，\ E(\Delta B_{t_k^{(n)}})^4 = 3(\Delta t_k^{(n)})^2，$$

$$E(I_n - \frac{t}{2})^2 = \frac{1}{4}\sum_k E[(\Delta B_{t_k^{(n)}})^2 - \Delta t_k^{(n)}]^2 = \frac{1}{4}\sum_k E[(\Delta B_{t_k^{(n)}})^4 - (\Delta t_k^{(n)})^2]$$

$$= \frac{1}{2}\sum_k (\Delta t_k^{(n)})^2 \le \frac{1}{2}\Delta_n \sum_k \Delta t_k^{(n)} = \frac{T}{2}\Delta_n \to 0，(n\to\infty) .$$

这说明

$$E\,|\,J_n - (\frac{1}{2}B_t^2 - \frac{t}{2})\,|^2 \to 0 .$$

用 Chebyshev 不等式，便得J_n概率收敛到$\frac{1}{2}(B_t^2 - t)$．

［注］以上论证实际上证明了以下的命题:

命题12．46

$$E\,|\sum_k (\Delta B_{t_k^{(n)}})^2 - t\,|^2 \to 0 .$$

类似的推理可以证明下面的命题.

命题12.46'

$$E\,|\sum_k f(B_{t_k^{(n)}})(\Delta B_{t_k^{(n)}})^2 - \int_0^T f(B_t)dt\,|^2 \to 0 .$$

［注］　如果我们在例 12.45 的积分和中，用 $B_{t_{k+1}^{(n)}}$ 代替$B_{t_k^{(n)}}$．并记

$$H_n = \sum_k B_{t_{k+1}^{(n)}} \Delta B_{t_k^{(n)}} .$$

那么

$$H_n = \frac{1}{2}\sum_k (2B_{t_{k+1}^{(n)}}{}^2 - 2B_{t_{k+1}^{(n)}} B_{t_k^{(n)}})$$

$$= \frac{1}{2}\sum_k [\Delta(B_{t_{k+1}^{(n)}}{}^2) + (\Delta B_{t_k^{(n)}})^2] = J_n + \sum_k (\Delta B_{t_k^{(n)}})^2 \text{.}$$

于是有

$$E\,|\,H_n - (\frac{1}{2}B_t^2 + \frac{t}{2})\,|^2 \to 0\text{.}$$

用 Chebyshev 不等式，立得 H_n 按概率收敛到 $\frac{1}{2}(B_t^2 + t)$.

特别地，我们有

$$E\,|\sum_k \frac{B_{t_k^{(n)}} + B_{t_{k+1}^{(n)}}}{2} \Delta B_{t_k^{(n)}} - \frac{1}{2}B_t^2\,|^2 \to 0\text{.}$$

[注] (B_t) 可知的随机过程 Φ_t 的可积条件(12．25)还可以减弱.

定义１２．４７　（Stratonovich(随机)积分）　若 f 为连续可微函数，则

$$\sum_k \frac{f(B_{t_k^{(n)}}) + f(B_{t_{k+1}^{(n)}})}{2} \Delta B_{t_k^{(n)}}$$

的概率极限，称为 Stratonowich 积分，记为 $\int_0^T f(B_t) \circ dB_t$. 又若 x_t 是一个 $(B_t) -$ 可知的连续随机过程，那么，类似地定义 $\int_0^T f(x_t) \circ dB_t$.

用与上面类似的推理，可以得到 Stratonowich 积分与 Ito 积分的如下关系：

$$\int_0^t f(B_s) \circ dB_s = \int_0^t f(B_s) dB_s + \frac{1}{2}\int_0^t f'(B_s) ds\text{.} \qquad (12．26)$$

(但是 $\int_0^t f(x_s) \circ dB_s \neq \int_0^t f(x_s) dB_s + \frac{1}{2}\int_0^t f'(x_s) ds$，在3.2段中，我们将给出其正确的形式).

例１２．４８

$$\int_0^t B_t \circ dB_t = \frac{1}{2}B_t^2\text{.}$$

例１２．４９　如果函数 f 有原函数 F，则用积分和可以直接验证

$$\int_0^t f(B_t) \circ dB_t = F(B_t) - F(B_0)\text{.}$$

这个结果与普通微积分中的结果是一致的，而 Ito 积分就没有这样简单的形式．这是 Stratonowich 积分的优点．但是 Ito-随机积分在计算中有其更为方便的优点，它便于利用鞅论作理论推导的工具．所以在很多情况下，数学家更喜欢使用 Ito 积分，而物理学家则更喜欢 Stratonovich 积分．

Ito 积分有下面的一些基本性质：

（1）线性性质

$$\int_0^t (\Phi+\Psi)dB = \int_0^t \Phi dB + \int_0^t \Psi dB,$$

$$\int_0^t (c\Phi)dB = c\int_0^t \Phi dB.$$

（对于一个只依赖于$\{B_u: u\leq a\}$的随机时间h，也有

$$\int_a^b (h\Phi)dB = h\int_a^b \Phi dB).$$

(2）可加性：对$a<b<c$有（对于$a<b$，定义

$$\int_a^b \Phi dB = \int_0^b \Phi dB - \int_0^a \Phi dB)$$

$$\int_a^c \Phi dB = \int_a^b \Phi dB + \int_b^c \Phi dB.$$

（3）对任意(B_t)有界停时τ，假定$t\leq T$，则可以定义

$$\int_0^t \Phi dB \overset{\Delta}{=} \int_0^T \Phi_t I_{(0,t]}(t)dB_t .$$

此外，作为“随机的”积分，它还有性质：

（4）零期望性与鞅性： 在t变化时，随机过程$\{\int_0^t \Phi_s dB_s\}$是(B_t)鞅 （这是 Brown 运动是鞅的推广），且

$$E(\int_0^t \Phi_s dB_s)=0 .$$

（5）协方差与平方可积鞅性质：对任意$0\leq u<t$有

$$E(\int_u^t \Phi_s dB_s \int_u^t \Psi_s ds)=E(\int_u^t \Phi_s\Psi_s ds)$$ （这是 Brown 运动平方可积鞅的推广），

从而随机过程$(\int_0^t \Phi_s dB_s)^2-\int_0^t \Phi_s^2 ds\}$也是$(B_t)$鞅.

((1),(2),(3)直接得自定义. 对(4)我们给出直观证明如下: 让我们简单地使用记号

$$\mathcal{F}_t \stackrel{\Delta}{=} \{B_u : u \le t\}.$$

对于 $s < t$,由随机过程 Φ_t 是 (B_t) 可知的,故

$$E\{\int_0^s \Phi_u dB_u \mid \mathcal{F}_s) = \int_0^s \Phi_u dB_u .$$

再则,对于 $s \le t_k < t_{k+1}$,我们有

$$E(\Phi_{t_k} \Delta B_{t_k} \mid \mathcal{F}_s) = E[E(\Phi_{t_k} \Delta B_{t_k} \mid \mathcal{F}_{t_k}) \mid \mathcal{F}_s] = E[\Phi_{t_k} E(\Delta B_{t_k} \mid \mathcal{F}_{t_k}) \mid \mathcal{F}_s] = 0 .$$

对于 k 求和,再取极限便得

$$E\{\int_s^t \Phi_u dB_u \mid \mathcal{F}_s) = 0 .$$

合起来就是

$$E\{\int_0^t \Phi_u dB_u \mid \mathcal{F}_s) = \int_0^s \Phi_u dB_u .$$

这就直观地证明了(4). 而这里的直观证明是利用了线性性质,Brown 运动的性质,条件期望的性质,最后还要加上极限与取条件期望的次序可以交换. 类似的考虑,可以得到(5)的直观证明.而性质(4),(5)的严格证明要涉及测度论的许多知识,本书从略.

*(6) 令

$$M_t = e^{\int_0^t \Phi_s dB_s - \frac{1}{2}\int_0^t \Phi_s^2 ds}, \tag{12.27}$$

则 M_t 是 (B_t) 鞅,称为 Ito 积分的指数鞅. 它是 Brown 运动的指数鞅的自然推广,我们将在下一段中用 Ito 公式证明它.

需要特别强调的是,性质(4)与(5)是使 Ito 积分比 Stratonovich 积分更易于用于随机分析的理论推导的原因.

3.2 Ito 公式—随机积分的换元公式与复合函数的随机微分公式

在普通函数微积分中,复合函数的微分公式与积分的换元公式是相互等价的两个最基本的公式. 而在随机微积分中,我们也有一个对应的公式,虽然它比微积分中的公式复杂,但是通过它仍然可以如在微积分中那样,把 Ito 积分方便地应用到许多问题中去.

定义12.50 (Ito 过程)

设

$$x_t = x + \int_0^t \Phi_s dB_s + \int_0^t \Psi_s ds ,$$

其中随机过程 Φ_t, Ψ_t 都是 (B_t) 可知的,且对任意固定的ω,都是 t 的连续函数, 满足

$\int_0^\infty E\Phi_s^2 ds < \infty$，$\int_0^\infty E|\Psi_s|ds < \infty$，而$\int_0^t \Psi_s ds$理解为在$w$固定后的普通积分，则$x_t$称为 Ito 过程．它也可以记成如下的 Ito 形式微分

$$dx_t = \Phi_s dB_s + \Psi_s ds.$$

设x_t是 Ito 过程．又二元实函数$f(t,x)$对x二阶光滑且对t一阶光滑．令h_t为复合得到的随机过程 $h_t = f(t,x_t)$， 则下面的 Ito 公式表明h_t也是一个 Ito 过程．也就是说，Ito 过程对于这种光滑函数的复合运算是封闭的．

我们先分析例 12.35， 它说明了$\frac{1}{2}B_t^2$是一个 Ito 过程，且 $d(\frac{1}{2}B_t^2) = B_t dB_t + dt$．于是对于 $f(x) = \frac{1}{2}x^2$，有 $df(B_t) = d(\frac{1}{2}B_t^2) \neq B_t dB_t = f'(B_t)dB_t$．可见为了得到 $df(B_t)$，仅用通常的微积分中的一阶展开 $f'(B_t)dB_t$ 是不够的，必须还补充以 $f(x) = \frac{1}{2}x^2$ 的 Taylor 展开中的第二项 $\frac{1}{2}f''(B_t)(dB_t)^2 = (dB_t)^2$，它提供了 $d(\frac{1}{2}B_t^2)$ 中的第二项 dt．比较后可见有下面的引理．

引理１２.５１

$$(dB_t)^2 = dt, \quad \text{即} \quad dB_t = (dt)^{\frac{1}{2}}. \qquad (12.28)$$

它说明dB_t是(dt 的)半阶无穷小$(dt)^{\frac{1}{2}}$． 从而对于 Ito 过程的复合函数$f(t,x_t)$，在作微分(即随机微分)的时候，必须把 Taylor 展开应用到二阶：

$$df(x_t) = f'(x_t)dx_t + \frac{1}{2}f''(x_t)(dx_t)^2, \qquad (12.29)$$

这才穷尽了半阶无穷小作出的的贡献．例１２．３５的结论可以一般化为下面的定理．

定理 12.2（Ito 公式，随机微分公式）

设x_t为 Ito 过程， 即$dx_t = \Phi_t dB_t + \Psi_t dt$． 二元实函数$f(t,x)$对$x$二阶光滑且对$t$一阶光滑． 那么，$h_t = f(t,x_t)$也是 Ito 过程，而且有

$$dh_t = df(t,x_t) + \frac{1}{2}d^2 f(t,x_t) = f_t'(t,x_t)dt + f_x'(t,x_t)dx_t + \frac{1}{2}f_{xx}''(t,x_t)(dx_t)^2$$

$$= (f_t' + f_x'\Psi_t + \frac{1}{2}\Phi_t^2 f_{xx}'')dt + f_x'\Phi_t dB_t, \qquad (12.30)$$

其中

$$(dx_t)^2 = (\Phi_t dB_t + \Psi_t dt)^2 \stackrel{\Delta}{=} \Phi_t^2 dt. \qquad (12.31)$$

(理解与证明 Ito 公式的核心是：$(dB_t)^2 = dt$，即 $dB_t = (dt)^{\frac{1}{2}}$,在严格的论证中它实际上

将被关系 $E|\sum_k(\Delta B_{t_k^{(n)}})^2-t|^2\to 0$ 所代替 ）.

Ito 公式实际上就是用 Taylor 公式对 $f(t,x_t)$ 作关于 dt 的一阶近似．它也就给出了 Ito 公式的直观推导．而其严格推导则需要较为精致的随机分析工具，本书中并不必要介绍．

推论１２.５２ 设 x_t 为 Ito 过程：$dx_t=\Phi_t dB_t+\Psi_t dt$． $f(x)$ 二阶光滑．那么，$f(x_t)$ 也是 Ito 过程，而且有

$$df(x_t)=df(x_t)+\frac{1}{2}d^2f(x_t)=f'(x_t)dx_t+\frac{1}{2}f''(x_t)(dx_t)^2$$
$$=(f'\Psi_t+\frac{1}{2}\Phi_t^2 f'')dt+f_x'\Phi_t dB_t \qquad (12.32)$$

注 1 设 x_t 为 Ito 过程：$dx_t=\Phi_t dB_t+\Psi_t dt$． $f(x)$ 二阶光滑．那么，

$$\int_0^t f(x_s)\circ dB_s=\int_0^t f(x_s)dB_s+\frac{1}{2}\int_0^t f'(x_s)\Phi_s ds$$

注 2 设 x_t 是 Stratonovich 型的 Ito 过程 $x_t=x+\int_0^t\Phi_s\circ dB_s+\int_0^t\Psi_s ds$ （写成 $dx_t=\Phi_t\circ dB_t+\Psi_t dt$），那么,可以证明它有如普通的微积分那样的复合函数的微分的链法则； 即对于三阶可微的函数 $f(x)$ 有 $df(x_t)=f'(x_t)\circ dx_t$.

定理１２．５３（多维 Ito 公式）

设 $\vec{x_t}=(x_t^{(1)},\cdots,x_t^{(d)})$， $x_t^{(i)},(i\le d)$ 为 Ito 过程：$dx_t^{(i)}=\Phi_t^{(i)}dB_t+\Psi_t^{(i)}dt$． $f(t,\vec{x})$ 为对 $\vec{x}$ 二阶光滑且对 t 一阶光滑．那么，$h_t=f(t,\vec{x_t})$ 也是 Ito 过程，而且有

$$dh_t=df(t,x_t)+\frac{1}{2}d^2f(t,x_t)$$
$$=f_t'(t,x_t)dt+\sum_{j=1}^d f_{x_i}'(t,\vec{x_t})dx_t^{(i)}+\frac{1}{2}\sum_{i,j=1}^d f_{x_ix_j}''(t,\vec{x_t})(dx_t^{(i)}dx_t^{(j)})$$
$$=[f_t'+\sum_{i=1}^d f_{x_i}'\Psi_t^{(i)}+\frac{1}{2}\sum_{i,j=1}^d\Phi_t^{(i)}\Phi_t^{(j)}f_{x_ix_j}'']dt+[\sum_{i=1}^d f_{x_i}'\Phi_t^{(i)}]dB_t, \qquad (12.33)$$

其中

$$dx_t^{(i)}dx_t^{(j)}=(\Phi_t^{(i)}dB_t+\Psi_t^{(i)}dt)(\Phi_t^{(j)}dB_t+\Psi_t^{(j)}dt)\stackrel{\Delta}{=}\Phi_t^{(i)}\Phi_t^{(j)}dt. \qquad (12.34)$$

推论１２．５４ （乘积公式） 设 x_t,h_t 都是 Ito 过程，则

$$d(x_t h_t) = x_t dh_t + h_t dx_t + (dx_t)(dh_t) . \qquad (12.35)$$

推论１２．５５（推广的 Ito 公式）

设x_t为 Ito 过程，$g(x)$为有界连续函数（显见$\int_0^t g(x_s)ds$也是 Ito 过程）．又若 $f(t,x)$ 是对 x 二阶光滑，且对 t 一阶光滑的实函数，$h(x)$ 一阶连续可微．那么，$h_t = h(\int_0^t g(x_s)ds) f(t,x_t)$ 也是 Ito 过程，而且有

$$dh_t == f(t,x_t)h'(\int_0^t g(x_s)ds)g(x_t)dt + h(\int_0^t g(x_s)ds)df(t,x_t) .$$

（证明　用二维过程$(x_t, \int_0^t g(x_s)ds)$的 Ito 公式）．

例１２．５６　设s ,b是常数，Gauss 过程 $h_t = e^{-bt}[h_0 + s\int_0^t e^{bs}dB_s]$，则有

$$dh_t = -bh_t dt + sdB_t .$$

证明　令 $x_t = h_0 + s\int_0^t e^{bs}dB_s$，则$h_t = e^{-bt}x_t$．对它用 Ito 公式便得到

$$dh_t = e^{-bt}dx_t - be^{-bt}x_t dt = s dB_t - bh_t dt$$

［注］这里h_t满足一个随机微分方程.(由它的定义还知道它是一个 Gauss 马氏过程，可以证明它就是参数为$b = b, g = \dfrac{s^2}{2b}$的 OU 过程.）　这个随机微分方程最早出现于理论物理中，称为 Langevin 方程．

例１２．５７　设Φ_t是有界的（ 即存在$M > 0$，使得对于任意(t,w)有 $|\Phi_t(w)| \le M$），且为(B_t)可知的随机过程．又

$$Z_t = e^{x_t}, x_t = \int_0^t \Phi_s dB_s - \frac{1}{2}\int_0^t \Phi_s^2 ds .$$

那么，由 Ito 公式推出

$$dZ_t = de^{x_t} = e^{x_t}dx_t + \frac{1}{2}e^{x_t}(dx_t)^2$$

$$= e^{x_t}(\Phi_t dB_t - \frac{1}{2}\Phi_t^2 dt) + \frac{1}{2}e^{x_t}\Phi_t^2 dt = Z_t \Phi_t dB_t \text{ .}$$

可见 Z_t 是随机微分方程

$$dZ_t = Z_t \Phi_t dB_t$$

满足初始条件

$$Z_0 = 1$$

的解．事实上由下面的定理12.59知道此解是唯一的．又由于 Z_t 作为 Ito 过程只含对 Brown 运动的随机积分项，由 Ito 积分的性质(４)得到 Z_t 是 (B_t) 鞅．这就顺便地证明了 Ito 积分的指数鞅性质．

例１２．５８（描述证券的 Black-Scholes 模型）

Black-Scholes 用如下的随机微分方程(称为 Black-Scholes 随机微分方程)

$$dx_t = x_t(bdt + s\,dB_t)$$

的解 x_t 来描述证券价格的随机模型，其中 $bdt + sdB_t$ 称为随机的收益变化率, 常数 b 称为平均收益率（Yield），常数 s 称为波动率(Volatility)．这个方程与例１２．５７中的方程非常类似．我们容易用 Ito 公式直接验证

$$x_t \overset{\Delta}{=} x_0 e^{(b-\frac{s^2}{2})t+sB_t}$$

是 Black-Scholes 随机微分方程的解．当 $x_0 = 1$ 时，这个解就是几何 Brown 运动．

从本章第 4 节中的定理，我们可以知道，在初始值 x_0 给定的条件下，例１２．５６，例１２.５７与例１２．５８三个例子中的解，都是唯一的．

Ito 公式成立的范围可以更广，但是需要用测度论的语言．我们这里只能就特殊的情况，把 Ito 公式的核心内容介绍给读者， 使读者能领略随机微积分的概要．要完全地，严格地懂得与掌握随机微积分，读者必须先掌握基本的测度论知识与方法．

4．随机微分方程与扩散过程简介

4．1　随机微分方程

在例１２．５６与例１２．５７中，我们已经给出了两个特殊的随机微分方程：$dh_t = -bh_t + s\,dB_t$ 与 $dZ_t = Z_t \Phi_t dB_t$ ．本段将介绍较为一般的随机微分方程．

随机微分方程也称随机积分方程，是表达相当广的一类连续时间、连续状态的 Markov 过程的一个重要而方便的工具．随机微分方程的一般形式为

$$dx_t = b(t,x_t)dt + s\,(t,x_t)dB_t \text{ .} \qquad (12.36)$$

它应该理解为其积分形式

$$x_t = x_0 + \int_0^t b(s,x_s)ds + \int_o^t s(s,x_s)dB_s,$$

其中第一个积分理解为对任意固定的ω后,对时间参数作普通函数的积分，而第二个积分则就是为 Ito 积分.

上述随机微分方程解的存在性和唯一性的研究，是随机微分方程的理论研究的一个基本问题．它不仅是随机分析的理论基础，而且对于许多应用问题也有重要意义．例如，在物理问题中归纳出的 Langevin 方程，从其存在唯一性，就可断定它就是例１２．３５中给出的那个 Gauss　Markov 过程．由此我们就可以知道它的统计特性.

下面我们给出随机微分方程的一个最简单的存在唯一性定理.

定理１２．５９　如果 s , b 满足：存在 $C>0$，使对于任意 t，一致地满足如下的 Lipschitz 条件

$$|s(t,x)-s(t,y)|+|b(t,x)-b(t,y)| \le C|x-y|,$$

那么，在初始值 x_0 给定的条件下(指整个随机变量给定)，随机微分方程(12．36)存在唯一的解 x_t (其含义为：如果有另一个解 x_t'，那么只要满足 $P(x_0' = x_0)=1$，就一定有 $P(x_t' = x_t, \forall t>0)=1$)．此解 x_t 可以由下面迭代程序的收敛极限得到：

$$x_t^{(0)} = x_0,$$

$$x_t^{(n+1)} = x_0 + \int_0^t b(s,x_s^{(n)})ds + \int_o^t s(s,x_s^{(n)})dB_s.$$

这时对任意 $T>0$ 有

$$\int_0^T E|x_t^{(n)} - x_t|^2 \, dt \xrightarrow{n\to\infty} 0.$$

(此极限式的成立，还可以进一步保证对任意 $e>0, T>0$ 有

$$P(\sup_{t\le T} |x_t^{(n)} - x_t| > e) \xrightarrow{n\to\infty} 0).$$

由此也可看出，x_t 只依赖于 x_0 及 Brown 运动 $\{B_t\}$ 在时刻 t 以前的值．特别地，如果 $x_0 = x$ (非随机)，则随机过程 x_t 是 (B_t) 可知的.

[注]　本定理也是一个近似算法．再则，保证存在唯一性的条件实际上可以减弱．但本书中不再论述.

例１２．５６与例１２．５８中的随机微分方程，都满足定理１２．５９的条件．所以在初始值给定的条件下，方程的解是存在(已经得到了解析表达式)且唯一的.例１２．５７中的随机微分方程的解，也是唯一的，但是需要另行证明，本书从略.

例１２．６０　(线性随机微分方程)　对于线性随机微分方程

$$dX_t = (bX_t + a)dt + (s \cdot X_t + c)dB_t,$$

可以仿照常微分方程的方法得到它的解．首先，将方程改写为

$$dX_t - (bX_t dt + s \cdot X_t dB_t) = adt + cdB_t.$$

它所对应的齐次线性随机微分方程

$$dY_t - (bY_t dt + s \cdot Y_t dB_t) = 0$$

是一个 Black-Scholes **随机微分方程，故其解为**

$$Y_t = Y_0 e^{(b - \frac{s^2}{2})t + sB_t}.$$

仿照常微分方程中恰当因子的方法，把其倒数Y_t^{-1}**乘到非齐次方程上，以便用** Ito **公式求得** $d(X_t Y_t^{-1})$**．由于．**$Y_t^{-1} = Y_0^{-1} e^{-(b - \frac{s^2}{2})t - sB_t} = Y_0^{-1} e^{[(-b + s^2) - \frac{s^2}{2}]t - sB_t}$，**可见**$Y_t^{-1}$**应满足以下方程**

$$dY_t^{-1} = Y^{-1}[(-b + s^2)dt - s \cdot dB_t]．$$

因此

$$d(X_t Y_t^{-1}) = XdY^{-1} + Y^{-1}dX + dXdY^{-1}$$

$$= Y^{-1}X[(-b + s^2)dt - s \cdot dB_t] + Y^{-1}[(bX + a)dt + (s \cdot X + c)dB] - (s \cdot X + c)Y^{-1}s \cdot dt$$

$$= Y^{-1}[(a - cs)dt + cdB].$$

于是就得到解 X_t **的显式表示**

$$X_t = Y_t[X_0 Y_0^{-1} + (a - cs)\int_0^t Y_s^{-1} ds + c\int_0^t Y_s^{-1} dB_s]$$

$$= X_0 e^{(b - \frac{s^2}{2})t + sB_t} + (a - cs)\int_0^t e^{(b - \frac{s^2}{2})(t-s) + s(B_t - B_s)} ds + c\int_0^t e^{(b - \frac{s^2}{2})((t-s) + s(B_t - B_s))} dB_s].$$

例１２．６１　求解

$$dX_t = (aX_t + b\sqrt{X_t})dt + cX_t dB_t.$$

我们可以通过变换，并运用 Ito **公式，把此方程化为一个可以视随机性**w**为参变量的普通的常微分方程，从而可以求得方程的显式解．此解为**

$$X_t = e^{cB_t - \frac{c^2}{2}t + at}(\frac{b}{2}\int_0^t e^{-\frac{c}{2}B_s + (\frac{c^2}{4} - \frac{a}{2})s} ds + \sqrt{X_0})^2.$$

(**证明提示　对** $dX_t = (aX_t + b\sqrt{X_t})dt + cX_t dB_t$，**令** $M_t = e^{-cB_t + \frac{1}{2}c^2 t}$，**则有**

$$d(M_t X_t) = M_t(aX_t + b\sqrt{X_t})dt.$$

记 $Y_t = M_t X_t$，那么它满足如下的常微分方程

$$\frac{dY_t}{dt} = aY_t + be^{-\frac{1}{2}cB_t + \frac{1}{4}c^2 t}\sqrt{Y_t}\ .$$

再作变换 $Z_t = \sqrt{Y_t}$， 则得到 $\frac{dZ_t}{dt} = \frac{a}{2}Z_t + \frac{b}{2}e^{-\frac{1}{2}cB_t + \frac{1}{4}c^2 t}$. 求解 Z_t 后再代回去便得.

例１２．６２ （时变 Kalman-Bucy 滤波的连续时间形式） 设 $\{B_t : t \geq 0\}, \{W_t : t \geq 0\}$ 是两个彼此独立的 Brown.运动，而状态随机过程 X_t 满足如下的线性随机微分方程

$$dX_t = a(t)X_t dt + b(t)dB_t\ .$$

但是状态 X_t 无法测量得到．能测量到的是与之有关的，如下的量测随机过程 Y_t

$$dY_t = c(t)X_t dt + d(t)dW_t\ ,$$

其中 $a(t), b(t), c(t), d(t)$ 都是非随机的函数，则用资料 $\{Y_s : s \leq t\}$ 得到 X_t 的估计 $\hat{X}_t$ 满足如下的方程，称为滤波方程（实际应用时，应该用差分采样代替，从而得到滤波公式）

$$d\hat{X}_t = a(t)\hat{X}_t dt + \frac{c(t)p(t)}{d(t)^2}[dY_t - c(t)\hat{X}_t dt]$$

$$p'(t) = 2a(t)p(t) - [\frac{c(t)}{d(t)}]^2 p^2(t) + b^2(t)\ .$$

这里 $p(t)$ 满足的方程是 Riccatti 方程，可以用变换 $p(t) = \frac{d^2(t)}{c(t)}\frac{u'(t)}{u(t)}$ 化为 $u(t)$ 的二阶线性方程以求解. (证明较长从略)．

４．２ 扩散过程

定理１２．６３ 在定理１２．５９的条件成立下，将随机微分方程

$$x_t = x + \int_0^t b(x_s)ds + \int_o^t s(x_s)dB_s$$

的唯一解记为 $x_t^{(x)}$ （因为它依赖初始值 x）．那么 $x_t^{(x)}$ 是具有概率转移密度的时齐 Markov 过程，其转移密度 $p(t, x.y)$ 存在，且满足下面的扩散方程(也称为 Kolmogorov 向后方程):

$$\frac{\partial p}{\partial t} = \frac{s(x)^2}{2}\frac{\partial^2 p}{\partial x^2} + b(x)\frac{\partial p}{\partial x}\ , \qquad (12．37)$$

$$\int p(0, x, y)f(y)dy = f(x) \quad \text{（初值）.} \qquad (12．38)$$

又因为转移密度 $p(t,x.y)$ 满足扩散方程，所以随机过程 $x_t^{(x)}$ 称为扩散过程． 而 $a(x)\overset{\Delta}{=}s\,(x)^2$ 称为扩散系数，$b(x)$ 称为漂移系数．

Markov 性的直观证明 对于任意 n，及任意 $t>s>s_n>\cdots>s_1$，在 $t\in[s,\infty)$，满足条件 $x_s=x,x_{s_n}=x_n,\cdots,x_{s_1}=x_1$ 的解 x_t 的迭代构造也应为

$$x_t^{(0)}=x\,,x_t^{(n+1)}=x+\int_s^t b(x_s^{(n)})ds+\int_s^t s\,(x_s^{(n)})dB_s\,,$$

其中的一切 $x_t^{(n)}$ 都只依赖于 $x,s\,(x),b(x)$，而不依赖于 $x_n,\cdots,x_1$．从而条件概率 $P(x_t\in\Lambda\,|\,x_s=x,x_{s_n}=x_n,\cdots,x_{s_1}=x_1)$ 也不依赖于 $x_n,\cdots,x_1$，即应该有

$$P(x_t\in\Lambda\,|\,x_s=x,x_{s_n}=x_n,\cdots,x_{s_1}=x_1)=P(x_t\in\Lambda\,|\,x_s=x)\,,$$

此即 Markov 性．

往证时齐性．对于方程的解有

$$x_{t+s}=x_s+\int_s^{t+s} b(x_u)du+\int_s^{t+s} s\,(x_u)dB_u\,.$$

改记

$$\overline{x}_t=x_{t+s}\,,\ \overline{B}_t=B_{t+s}-B_s\,,$$

则 $\overline{B}_t$ 也是一个 Brown 运动．而且直接由 Ito 积分的定义，从极限容易得到

$$\int_s^{t+s} s\,(x_u)dB_u=\int_0^t s\,(\overline{x}_u)d\overline{B}_u\,.$$

于是

$$\overline{x}_t=\overline{x}_0+\int_0^t b(\overline{x}_u)du+\int_0^t s\,(\overline{x}_u)d\overline{B}_u\,.$$

与

$$x_t=x_0+\int_0^t b(x_u)du+\int_0^t s\,(x_u)dB_u$$

是具有相同系数但对应于不同 Brown 运动的 Ito 方程的解．而对于不同的 Brown 运动对应的迭代过程的每一步都具有相同的分布，从而它们的解作为极限有相同的转移函数．

关于转移密度的存在性证明，一般都较为复杂．最方便的是援用光滑系数的线性偏微分方程的结论：在 $s^{\,2}(x),b(x)$ 有高于二阶的连续导数时，向后方程

$$\frac{\partial u(t,x)}{\partial t}=\frac{1}{2}s^{2}(x)\frac{\partial^{2}u}{\partial x^{2}}+b(x)\frac{\partial u}{\partial t}$$

的基本解$G(t,x,y)$对t一阶连续可导，对x二阶连续可导，而且$u(t,x)=\int G(t,x,y)u(0,y)dy$，

此时$G(t,x,y)$正是我们的转移密度$p(t,x,y)$.

对于 Kolmogorov 向后方程，我们将在定理 12.63 后面的注中给出其直观推导.

［注］由

$$x_t-x_s=\int_s^t b(x_u)du+\int_s^t s(x_u)dB_u$$

可得漂移系数与扩散系数的概率含义为

$$b(x)=\lim_{h\to 0}\frac{E[(x_{s+h}-x_s)\,|\,x_s=x]}{h}$$ （即条件平均速率），

$$a(x)=\lim_{h\to 0}\frac{E[(x_{s+h}-x_s)^2\,|\,x_s=x]}{h}$$ （即条件平均二阶矩增长率）.

定理１２．６３ （Kolmogorov 向前方程，Fokker-Plank 方程）

若$s(x)$二阶光滑，$b(x)$一阶光滑，则扩散过程的转移密度$p(t,x,y)$还满足如下的方程

$$\frac{\partial p}{\partial t}=\frac{\partial^{2}}{\partial y^{2}}\left(\frac{s(y)^{2}}{2}p\right)-\frac{\partial}{\partial y}(b(y)p),\qquad (12.39)$$

它称为 Kolmogorov 向前方程 或 Fokker-Plank 方程．写成散度型的方程就是

$$\frac{\partial p}{\partial t}=\frac{\partial}{\partial y}\left(\frac{s(y)^{2}}{2}\frac{\partial p}{\partial y}\right)-\frac{\partial}{\partial y}([b(y)-s(y)s'(y)]p).\qquad (12.39)'$$

［注］Kolmogorov 向前方程与向后方程的直观推导：

（1）Kolmogorov 向前方程的直观推导． 由 Ito 公式有

$$f(x_t)=f'(x_t)dx_t+\frac{1}{2}f''(x_t)(dx_t)^2,$$

$$f(x_{t+s})=f(x_s)+\int_s^{s+t}\left[\frac{1}{2}s^{2}(x_u)+b(x_u)\right]f'(x_u)dt+\int_s^{s+t}s(x_u)f''(x_u)dB_t$$

对于有连续二阶导数且在有限区间外取值 0 的函数(简称为C_0^2函数)$f(x)$而言，$\int_0^t b(x_u)f(x_u)dB_u$

是(B_t)鞅， 记$\mathcal{F}_t=\{B_u:u\le t\}$，由条件期望的性质得到

$$E(f(x_{t+s})\,|\,x_0=x)-E(f(x_t)\,|\,x_0=x)=E((E[(f(x_{s+t})-f(x_t))\,|\,\mathcal{F}_t]\,|\,x_0=x)$$

$$= E[(E(\int_t^{s+t} [\frac{1}{2} s^2(x_u) f''(x_u) + b(x_u) f'(x_u)]du) \mid x_t) \mid x_0 = x].$$

再用微积分中的定理，我们有

$$\frac{1}{s}(E([f(x_{t+s}) - f(x_t)] \mid x_0 = x)$$

$$= E(\frac{1}{s}\int_t^{s+t} E[(\frac{1}{2} s^2(x_u) f''(x_u) + b(x_u) f'(x_u)) \mid x_t] du \mid x_0 = x)$$

$$\xrightarrow{s \to 0} E[(\frac{1}{2} s^2(x_t) f''(x_t) + b(x_t) f'(x_t)) \mid x_0 = x].$$

另一方面，如果我们假定时齐的 Markov 过程 x_t 有转移密度 $p(t, x, y)$，那么由

$$E(f(x_t) \mid x_0 = x) = \int p(t, x, y) f(y) dy,$$

上面的极限式就变成

$$\frac{1}{s}\int (p(t+s, x, y) - p(t, x, y)) f(y) dy$$

$$\xrightarrow{s \to 0} \int p(t, x, y)[\frac{1}{2} s^2(y) f''(y) + b(y) f'(y)] dy.$$

即

$$\int \frac{\partial p(t, x, y)}{\partial t} f(y) dy = \int p(t, x, y)[\frac{1}{2} s^2(y) f''(y) + b(y) f'(y)] dy.$$

再用分部积分推出(利用 f 在 $\pm\infty$ 处为 0)

$$\int \frac{\partial p(t, x, y)}{\partial t} f(y) dy = \int [\frac{1}{2} \frac{\partial^2}{\partial y^2}(s^2(y) p(t, x, y)) - \frac{\partial}{\partial y}(b(y) p(t, x, y))] f(y) dy.$$

由于 f 的任意性，我们便得到

$$\frac{\partial p(t, x, y)}{\partial t} = \frac{1}{2} \frac{\partial^2}{\partial y^2}(s^2(y) p(t, x, y)) - \frac{\partial}{\partial y}(b(y) p(t, x, y)).$$

这正是 Kolmogorov 向前方程 (12. 39).

(2) Kolmogorov 向后方程的直观推导　对于任意 C_0^2 函数 $g(x)$，记

$$v(t, y) = \int p(t, x, y) g(x) dx.$$

则我们仍有

$$\frac{\partial v(t, y)}{\partial t} = \frac{1}{2} \frac{\partial^2}{\partial y^2}(s^2(y) v(t, y)) - \frac{\partial}{\partial y}(b(y) v(t, y)),$$

而且由 Chapman-Kolmogorov 方程，还有

$$v(t+s,y)=\int p(t+s,x,y)g(x)dx=\int\int p(t,x,z)p(s,z,y)dzg(x)dx$$
$$=\int p(s,z,y)v(t,z)dz=\int p(s,x,y)v(t,x)dx(=\int p(t,x,y)v(s,x)dx)$$

于是由上式及 Kolmogorov 向前方程和形式的分部积分(指终端$\pm\infty$处都认为取 0)推出

$$\int \frac{\partial p(t+s,x,y)}{\partial t}g(x)dx=\frac{\partial v(t+s,y)}{\partial t}=\frac{\partial v(t+s,y)}{\partial s}=\int p(t,x,y)\frac{\partial v(s,x)}{\partial s}dx$$

$$=\int p(t,x,y)(\frac{1}{2}\frac{\partial^2[s^2(x)v(s,x)]}{\partial x^2}-\frac{\partial[b(x)v(s,x)]}{\partial x})dx$$

$$=\int v(s,x)[\frac{1}{2}s^2(x)\frac{\partial^2 p(t,x,y)}{\partial x^2}+b(x)\frac{\partial p(t,x,y)}{\partial x}]dx$$

$$=\int\int p(s,y,x)g(y)dy[\frac{1}{2}s^2(x)\frac{\partial^2 p(t,x,y)}{\partial x^2}+b(x)\frac{\partial p(t,x,y)}{\partial x}]dx.$$

令 $s\to 0$ 就得到

$$\int \frac{\partial p(t,x,y)}{\partial t}g(x)dx=\int g(x)[\frac{1}{2}s^2(x)\frac{\partial^2 p(t,x,y)}{\partial x^2}+b(x)\frac{\partial p(t,x,y)}{\partial x}]dx.$$

再由 $g(x)$ 的任意性得到

$$\frac{\partial p(t,x,y)}{\partial t}=\frac{1}{2}s^2(x)\frac{\partial^2 p(t,x,y)}{\partial x^2}+b(x)\frac{\partial p(t,x,y)}{\partial x},$$

这正是 Kolmogorov 向后方程(12. 37).

需要指出的是，以上均为形式运算，对于函数的定义域，对于积分，微商，极限等诸多运算之间的可交换性，都未追究．所以我们的推导只是给出了猜测结果的一种方法．正式的数学证明，则需要使用较多的知识与相对地较为长的叙述.

定理１２．６４（Master 方程） 设定理１２．６３的条件成立，且x_0有分布密度（此条件非本质）．把 Ito 方程的解在时刻t的分布函数$P(x_t\le y)$的密度函数(它必定存在，但是这一事实并不容易证明)记为 $p(t,y)$，那么它满足以下的 Master 方程:

$$\frac{\partial p}{\partial t}=\frac{\partial^2}{\partial y^2}(\frac{s(y)^2}{2}p)-\frac{\partial}{\partial x}(b(y)p)\,. \qquad (12．40)$$

证明　设x_0的密度为 $p(0,y)$．因为 $p(t,y)=\int p(t,x,y)p(0,x)dx$，由转移密度满足的 Kolmogorov 向前方程便得．

.　下面我们讨论扩散过程$\{x_t,t\ge 0\}$的转移函数 $p(t,x,y)$ 的不变密度．一般并非所有的扩散过程的转移密度都有不变密度．例１２．９ 中，Brown 运动作为最简单的扩散过程，就不存在不变密度．对于扩散过程，有一个几乎公认的断言，就是对于较好的系数

$s(x), b(x)$ (例如，满足定理１２.６３的条件的情形)，作为随机微分方程解的扩散过程的转移密度具有如下的二岐性质：　在$t\to\infty$时，要么　$p(t,x,y)\to 0$ (例如 Brown 运动)，要么 $p(t,x,y)\to j(y)$，与x无关且$\int j(y)dy=1$．　这两种情形分别类似于 Markov 链的零常返与正常返．可是这个断言还未见严格的概率论方法的论证．在第二种情形，根据偏微分方程的理论，极限函数j　应该是 Kolmogorov 向前方程的定态解(即不依赖时间t的解)，即满足椭圆型方程

$$\frac{\partial^2}{\partial y^2}\left(\frac{s(y)^2}{2}j\right)-\frac{\partial}{\partial y}(b(y)j)=0\ . \qquad (12.\ 41)$$

另一方面，在 Chapman-Kolmogorov 方程　$p(t+s,x,y)=\int p(t,x,z)p(s,z,y)dz$ 中，令 $t\to\infty$，便得$j(y)=\int j(z)p(s,z,y)dz$．这说明了极限函数j 是扩散过程的转移函数的不变密度.

一般地，我们有

定理１２．６５　转移函数$p(t,x,y)$有不变密度的充要条件为：方程

$$\frac{d^2}{dy^2}\left(\frac{s(y)^2}{2}y\right)-\frac{d}{dy}(b(y)y)=0$$

有非负可积解．在条件成立时，$j\overset{\Delta}{=}\dfrac{y}{\int y}$就是扩散过程的不变密度．

注　对于 Stratonovich 型随机微分方程

$$x_t=x+\int_0^t b(x_s)ds+\int_o^t s(x_s)\circ dB_s$$

的解，可以化归 Ito 型随机微分方程

$$x_t=x+\int_0^t [b(x_s)+\frac{1}{2}s(x_s)s'(x_s)]ds+\int_o^t s(x_s)dB_s$$

处理，其 Kolmogorov 向前方程（Fokker-Planck 方程）则是如下的散度型的

$$\frac{\partial p}{\partial t}=\frac{\partial}{\partial y}\left(\frac{s(y)^2}{2}\frac{\partial p}{\partial y}\right)-\frac{\partial}{\partial y}\left([b(y)-\frac{1}{2}s(y)s'(y)]p\right)\ . \qquad (12.\ 39)''$$

警告　有一些书上将(12. 39)”误写为

$$\frac{\partial p}{\partial t}=\frac{\partial}{\partial y}\left(\frac{s(y)^2}{2}\frac{\partial p}{\partial y}\right)-\frac{\partial}{\partial y}(b(y)p)\ ,$$

这是不对的.

再则，如果我们定义如下的 新的随机积分：设 f 为连续可微函数，则当$n\to\infty$时

$$\sum_k f(x_{t_{k+1}^{(n)}})\Delta B_{t_k^{(n)}}$$

的概率极限，我们称之为 "向后随机积分"，并将它记为 $\int_0^T f(B_t)*dB_t$. 那么，对于由这种向后随机积分定义的随机微分方程

$$x_t = x + \int_0^t b(x_s)ds + \int_o^t s(x_s)*dB_s,$$

我们可以类似地得到一个扩散过程，其散度型的 Kolmogorov 向前方程（Fokker-Planck 方程）正是

$$\frac{\partial p}{\partial t} = \frac{\partial}{\partial y}\left(\frac{s(y)^2}{2}\frac{\partial p}{\partial y}\right) - \frac{\partial}{\partial y}(b(y)p).$$

而散度型的优点是，其二阶导数部分写成为一个对称的运算形式，在二阶线性抛物型偏微分方程中，这种形式是较易处理的.

总之，对于随机积分我们可有三种不同定义，它们各有优缺点： Ito 积分(向前积分)的优点是有鞅性质，Strtonovich 积分(中间随机积分)的优点是有复合函数的链法则，向后积分的优点是 Fokker-Planck 方程的散度型具有简单而直接的形式.

以上的有关扩散过程的结论，都可以推广到高维情形．对于 m 维 Brown 运动 $\vec{B}_t$，即 m 个相互独立的 Brown 运动排成的列向量，以及 $(\vec{B}_t)$ 可知的 $d\times m$ 矩阵值随机过程 Φ_t，可以类似地定义 Ito 积分和随机微分方程：对于 $d\times m$ 矩阵值函数 $\Sigma(x)$ 及 d 维向量值函数 $\vec{b}(x)$，d 维随机微分方程

$$\vec{x}_t = x + \int_0^t \vec{b}(\vec{x}_s)ds + \int_o^t \Sigma(\vec{x}_s)d\vec{B}_s.$$

有类似的存在唯一性定理．也有与例１２．５６至例１２．５８相应的方程的解的表达式．

定理１２.６６ d 维随机微分方程的解仍然是 Markov 过程，称为 d 维扩散过程，其转移密度存在,且满足以下诸方程：

Kolmogorov 向后方程：

$$\frac{\partial p}{\partial t} = \frac{1}{2}\sum_{i,j=1}^d a_{ij}(x)\frac{\partial^2 p}{\partial x_i \partial x_j} + \vec{b}(x)\cdot\nabla_x p,$$

其中 $x, y\in R^d$，$(a_{ij.}(x))_{i,j\le d} = \Sigma(x)\Sigma(x)^T$，$\nabla_x$ 是关于变量 x 的梯度运算．

Kolmogorov 向前方程：

$$\frac{\partial p}{\partial t} = \frac{1}{2}\sum_{i,j=1}^d \frac{\partial^2}{\partial y_i \partial y_j}(a_{ij}(y)p) - \nabla\cdot(\vec{b}(y)p),$$

其中 $\nabla\cdot$ 是对于变量 y 的散度运算．

扩散过程$\vec{x}_t$在时刻t的分布密度$p(t,x)$还满足 Master 方程：

$$\frac{\partial p}{\partial t}=\frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial^2}{\partial y_i \partial y_j}(a_{ij}(y)p)-\nabla\cdot(\vec{b}(y)p) .$$

从而其不变密度j 满足的方程为：

$$\frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial^2}{\partial y_i \partial y_j}(a_{ij}(y)j)-\nabla\cdot(\vec{b}(y)j)=0 . \qquad (12.41)'$$

［注］对于 Stratonovich 型随机微分方程

$$\vec{x}_t=\vec{x}+\int_0^t \vec{b}(\vec{x}_s)ds+\int_o^t \Sigma(\vec{x}_s)\circ d\vec{B}_s ,$$

其 Kolmogorov 向后方程为

$$\frac{\partial p}{\partial t}=\frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial^2}{\partial y_i \partial y_j}(a_{ij}(x)p)+\sum_{j=1}^{d}(b_j(y)+\frac{1}{4}\sum_{i=1}^{d}\frac{\partial a_{ij}}{\partial y_i})\frac{\partial p}{\partial y_j} .$$

而 Kolmogorov 向前方程可以写为以下的散度形式：
：

$$\frac{\partial p}{\partial t}=\frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial}{\partial y_i}(a_{ij}(y)\frac{\partial p}{\partial y_j})-\sum_{j=1}^{d}\frac{\partial}{\partial y_j}[(b_j(y)-\frac{1}{4}\sum_{i=1}^{d}\frac{\partial a_{ij}}{\partial y_i})p],$$

4. 扩散过程的遍历定理

由于扩散过程的遍历定理在物理,化学,生物,工程,经济,金融等诸多领域有广泛的应用,我们不加证明地叙述以下 3 个定理.

定理１２．６７ （遍历定理）如果存在$0<g_1<g_2$，使扩散矩阵$(A(x)=(a_{ij}(x))$满足

$$g_1 I\le A(x)\le g_2 I ，（I \text{ 是单位矩阵}），$$

而且方程

$$\frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial^2}{\partial y_i \partial y_j}(a_{ij}(y)j)-\nabla\cdot(\vec{b}(y)j)=0$$

有一个可积的正解．那么其归一化的正解$j(x)$是对应的扩散过程x_t的不变密度，并且成立如下的遍历定理：对于任意的有界的 Borel 函数f，不管x_0的初始分布是什么，恒有

$$P(\frac{1}{T}\int_0^T f(x_t)dt\xrightarrow{T\to\infty}\int f(x)j(x)dx)=1 .$$

(定理的证明远超出本书的水平．故而从略．这个定理是 Markov 链的遍历定理的扩散过程版本，其出处可参见下书中的附录 IV:

钱敏平，龚光鲁　随机过程论，第二版 ,北京大学出版社，1997，

该书中叙述的遍历性条件，是由 Orey 的二岐性定理保证成立的）.

类似的事实对于多个时刻也成立，即对于任意n，任意$0 < s_1 < \cdots < s_n$及任意的有界 Borel 函数$f(x_1, \cdots, x_n)$，不管x_0的初始分布是什么，总以概率为 1 地有

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T f(x_{t+s_1}, \cdots, x_{t+s_n})dt =$$

$$\int f(x_1, \cdots, x_n) j(x) p(s_2 - s_1, x_1, x_2) \cdots p(s_n - s_{n-1}, x_{n-1}, x_n) dx_1 \cdots dx_n .$$

作为定理１２.６７的推论，我们有

定理１２.６８　（平均遍历定理）

在定理１２。６７　的条件下，对于任意的有界的 Borel 函数f，及$x_0 = x$

$$\frac{1}{T}\int_0^T E_x f(x_t) dt \xrightarrow{T\to\infty} \int f(y) j(y) dy .$$

证明从略.

*4．3　Girsanov 定理与 Feyman-Kac 公式

Girsanov 定理与 Feyman-Kac 公式是随机微积分中，除了 Ito 公式以外最重要的两个定理．它们与 Ito 公式一起，形成了随机分析理论与应用的三大支柱.它们的证明需要随机分析的工具，所以我们只给出叙述，而不给证明.

定义１２．６８（Girsanov 变换）

设$\vec{B}_t$为概率空间$(\Omega, \mathcal{F}, P)$上的d维 Brown 运动，$\vec{\Phi}_t$为$(\vec{B}_t)$可知的有界随机过程．令

$$\hat{B}_t = \vec{B}_t - \int_0^t \vec{\Phi}_s ds ,$$

它称为 Brown 运动的随机平移或 Girsanov 变换．　定义新的概率P^*如下：　对于$\mathcal{F}$中的任意事件A，只要它的信息完全可由$\{B_s : s \le t)$确定，就定义它的新的概率为

$$P^*(A) = E[I_A e^{\int_0^t \vec{\Phi}_s^T d\vec{B}_s - \frac{1}{2}\int_0^t \Phi_s^T \Phi_s ds}],$$

则不难证明P^*确是$\mathcal{F}$上的概率.

定理１２．６９　（Girsanov 定理）　在新概率P^*下，Brown 运动的 Girsanov 变换$\hat{B}_t$是d维 Brown 运动.（警告：在原来的概率P下它并不是 **Brown** 运动）.

Girsanov 定理有广泛的应用．例如，把它应用到金融数学中的 Black-Scholes 模型，就可以得到风险中性的概率，这是期权(option)定价的理论基础.

定理１２.７０　（**Feyman-Kac** 公式）

设 $\vec{x}_t$ 是随机微分方程 $\vec{x}_t = x + \int_0^t \vec{b}(\vec{x}_s)ds + \int_o^t \Sigma(\vec{x}_s)d\vec{B}_s$ 的解，$c(x) \ge c_0 > -\infty$，

$(a_{ij.}(x))_{i,j\le d} = \Sigma(x)\Sigma(x)^T$．那么，在 $0 \le t \le T$ 满足终值条件

$$u(T,x) = f(x), (x \in R^d)$$

的偏微分方程

$$\frac{\partial u}{\partial t} + \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}(x)\frac{\partial^2 u}{\partial x_i \partial x_j} + \vec{b}(x)\nabla_x u - c(x)u + g(t,x) = 0,$$

的解可以用概率表示为以下的条件期望的形式：

$$u(t,x) = E([\int_t^T e^{-\int_t^s c(\vec{x}_u)du} g(s,\vec{x}_s) + e^{-\int_t^T c(\vec{x}_u)du} f(\vec{x}_T)] \mid \vec{B}_v : v \le t).$$

Feyman-Kac 公式把一个线性偏微分方程的终值问题的解，表达为一个随机微分方程的解的函数的条件期望．于是我们可以通过随机模拟得到线性偏微分方程的终值问题的数值解．例如说，得到

$$u(0,x) = E[\int_t^T e^{-\int_t^s c(\vec{x}_u)du} g(s,\vec{x}_s) + e^{-\int_t^T c(\vec{x}_u)du} f(\vec{x}_T)]$$

的近似值．

5　随机微分方程的解的数值模拟算法

对于随机微分方程

$$x_t = x_0 + \int_0^t b(s,x_s)ds + \int_o^t s(s,x_s)dB_s,$$

可以用迭代方法，或与常微分方程类似地，给出数值解的方法．但是，这样的方法在计算上是不经济的，需要采用更为实用的算法．

5．1　随机微分方程在固定时刻附近的随机 Taylor 展开与解的差分近似

在 $[0,T]$ 上随机微分方程

$$dX_t = b(X_t)dt + s(X_t)dB_t,\quad X_0 = x$$

的解是一个 Ito 过程. 求方程的数值解实际上是给出该过程的样本在离散时间采样点上的 Monte Carlo 近似. 为此, 我们先将 $[0,T]$ 作 n 等分

$$0 = t_0 < t_1 < \cdots < t_n = T,$$

使步长 $d = \frac{T}{n}$ 适当地小.

假定扩散系数 $s(x)$ 有界且二次连续可微. 我们考虑 X_t 在 $t_m\ (m < n)$ 附近的展开, 其思

想是：用随机的差分来近似．在$t_m < s < t_{m+1}$，对$s(X_s)$用 Ito 公式得到

$$s(X_s) - s(t_m) = \int_{t_m}^{s} [b(X_u)s'(X_u) + \frac{1}{2}s^2(X_u)s''(X_u)]du + \int_{t_m}^{s} s(X_u)s'(X_u)dB_u .$$

于是有

$$X_{t_{m+1}} - X_{t_m} = \int_{t_m}^{t_{m+1}} b(X_s)ds + \int_{t_m}^{t_{m+1}} s(X_s)dB_s$$

$$= (b(X_{t_m})d + o(d)) + s(X_{t_m})(B_{t_{m+1}} - B_{t_m}) + \int_{t_m}^{t_{m+1}} [s(X_s) - s(X_{t_m})]dB_s \qquad (12.42)$$

$$= b(X_{t_m})d + s(X_{t_m})V_m + \int_{t_m}^{t_{m+1}}\int_{t_m}^{s} s(X_u)s'(X_u)]dB_u dB_s + o(d), \qquad (12.43)$$

其中$\{V_m\}$独立同分布且$V_m = B_{t_{m+1}} - B_{t_m} \sim N(0,d)$.

首先考虑（１２．４２）．由于

$$E[\int_{t_m}^{t_{m+1}} [s(X_s) - s(X_{t_m})]dB_s]^2 \le \max_{t_m < s < t_{m+1}} E[s(X_s) - s(X_{t_m})]^2 \Delta t_m = o(d),$$

粗略地有

$$\int_{t_m}^{t_{m+1}} [s(X_s) - s(X_{t_m})]dB_s = o(\sqrt{d}).$$

所以，只要$s(x)$有界连续(并不需要二次连续可微)，我们就由（１２．４２）得到随机微分

方程的精确到d的$\frac{1}{2}$阶（即精度为$d^{\frac{1}{2}}$）的如下差分近似．

$\frac{1}{2}$阶差分近似模型（Euler-Maruyama 近似）：　令

$$X_{t_{m+1}}^{(n)} = X_{t_m}^{(n)} + b(X_{t_m}^{(n)})\Delta t_m + s(X_{t_m}^{(n)})V_m, \quad X_{t_0}^{(n)} = x, \qquad (12.44)$$

其中$\{V_m\}$独立同分布且 $V_m = \Delta B_{t_m} \sim N(0,d)$ （注意$\Delta t_m = \frac{T}{n} = d$）．

较高精度的近似是利用（１２．４３）．注意到

$$\int_{t_m}^{t_{m+1}}\int_{t_m}^{s} s(X_u)s'(X_u)]dB_u dB_s = s(X_{t_m})s'(X_{t_m})\int_{t_m}^{t_{m+1}}\int_{t_m}^{s} dB_u dB_s + o(d)$$

$$= s(X_{t_m})s'(X_{t_m})\int_{t_m}^{t_{m+1}} (B_s - B_{t_m})dB_s + o(d)$$

$$= \sigma(X_{t_m})\sigma'(X_{t_m})[\frac{1}{2}(B_{t_{m+1}}^2 - B_{t_m}^2 - \Delta t_m) - B_{t_m}\Delta B_{t_m}] + o(\delta)$$

$$= \frac{1}{2}\sigma(X_{t_m})\sigma'(X_{t_m})[(\Delta B_{t_m})^2 - \Delta t_m] + o(\delta). \qquad (12.45)$$

由(12. 43)和(12. 45), 我们得到随机微分方程的精确到δ 的１次方的如下的差分近似．

一阶差分近似模型 (Milstein 近似)：

$$X_{t_{m+1}}^{(n)} = X_{t_m}^{(n)} + b(X_{t_m}^{(n)})\Delta t_m + \sigma(X_{t_m}^{(n)})V_m + \frac{1}{2}\sigma(X_{t_m}^{(n)})\sigma'(X_{t_m}^{(n)})(V_m^2 - \Delta t_m), \quad (12.46)$$

其中$\{V_m\}$独立同分布, $V_m \sim N(0,\delta)$, $X_{t_0}^{(n)} = x$.

具体做法是: 对于给定的n, 通过独立地生成的$N(0,\delta)$随机数$V_m,(m<n)$, 逐步地解出的 $X_{t_1}^{(n)},\cdots,X_{t_n}^{(n)}$ ，就得到随机微分方程的近似解在离散时间集合$\{t_1,\cdots,t_n\}$上的采样值. 用它们做折线内插近似, 或样条函数(spline)内插近似, 就得到随机微分方程的一个近似模拟轨道.

如果漂移系数$b(x)$与扩散系数有更多阶的连续导数, 我们还可以沿着这个思路构造$\frac{3}{2}$阶, 2 阶, 以至更高阶的差分近似模型. 但是, 对于通常的应用, 一阶近似模型已经够了．

下面的内容, 有助于我们对随机差分模型近似的认识.

５．２　Ito 过程的一个光滑函数f复合在时刻t附近的随机 Taylor 展开

(本段内容可参考

Platen, E. and Wagner, W. On a Taylor formula for a class of Ito process. Probability Math. Statistics 3, 37-51, 1982).

对于 Ito 过程X_t， 我们考虑$f(X_t)$在t附近的近似计算．为了求得较好的近似, 我们要在$t_m < t \le t + \Delta t_m$时用 Ito 过程作随机 Taylor 展开式, 它是 5．1 段的思路的推广. 因为 5．1 段可以看成$f(x) = x$的情形.

由 Ito 公式

$$f(X_t) = f(X_{t_m}) + \int_{t_m}^{t} [b(X_s)f'(X_s) + \frac{1}{2}\sigma^2(X_s)f''(X_s)]ds + \int_{t_m}^{t} \sigma(X_s)f'(X_s)dB_s$$

, 它的$\frac{1}{2}$阶随机 Taylor 展开式则是在上式中，将被积的随机过程X_s简单地用（它在t_m处的随机变量）X_{t_m}替代后, 再把差别统统放到一个余项中, 这就是

$$f(X_t) = f(X_{t_m}) + [b(X_{t_m})f'(X_{t_m}) + \frac{1}{2}\sigma^2(X_{t_m})f''(X_{t_m})]\int_{t_m}^{t} ds + \sigma(X_{t_m})f'(X_{t_m})\int_{t_m}^{t} dB_s + R$$

其中R是一个随机的余项.

较高精确度的是一阶随机 Taylor 展开，　为了得到它，我们把上式改写为

$$f(X_t)=f(X_{t_m})+[b(X_{t_m})f'(X_{t_m})+\frac{1}{2}s^{2}(X_{t_m})f''(X_{t_m})]\int_{t_m}^{t}ds+s(X_{t_m})f'(X_{t_m})\int_{t_m}^{t}dB_s$$

$$+\int_{t_m}^{t}[s(X_s)f'(X_s)-s(X_{t_m})f'(X_{t_m})]dB_s+R_1.$$

对于乘积复合函数$(s\cdot f')(X_s)$应用 Ito 公式，我们发现$s(X_s)f'(X_s)-s(X_{t_m})f'(X_{t_m})$中含 Ito 积分的项只有$\int_{t_m}^{s} s(X_u)(s\cdot f')'(X_u)dB_u$，它的$\frac{1}{2}$阶近似为 $s(X_{t_m})(s\cdot f')'(X_{t_m})(B_s-B_{t_m})$，其它部分都高于$\frac{1}{2}$阶． 于是在只考虑一阶近似时，我们把展开式写为

$$f(X_t)=f(X_{t_m})+[b(X_{t_m})f'(X_{t_m})+\frac{1}{2}s^{2}(X_{t_m})f''(X_{t_m})]\int_{t_m}^{t}ds+s(X_{t_m})f'(X_{t_m})\int_{t_m}^{t}dB_s$$

$$+s(X_{t_m})(s\cdot f')'(X_{t_m})]\int_{t_m}^{t}\int_{t_m}^{s}dB_u dB_s+R_1.. \qquad (12.\ 47)$$

这就是一阶随机 Taylor 展开式.

5. 3. **差分近似模型的改进**

用一阶随机 Taylor 展开的思想作差分近似的缺点，是需要计算扩散系数$s(x)$的微商. 借用常微分方程近似计算的思想，我们也有

一阶随机 Runge-Kutta 模型

$$X_{t_{m+1}}^{(n)}=X_{t_m}^{(n)}+b(X_{t_m}^{(n)})\Delta t_m+s(X_{t_m}^{(n)})V_m+\frac{1}{2}\{\frac{1}{\sqrt{d}}[s(X_{t_m}^{(n)}+s(X_{t_m}^{(n)})\sqrt{d})-s(X_{t_m}^{(n)})]\}(V_m^2-\Delta t_m),$$

$\{V_m\}$独立同分布，$V_m\sim N(0,d)$， $X_{t_0}^{(n)}=x$． (12. 48)

当$n\to\infty$时，可以在理论上证明，以上几个模型得到的”近似解”确实是在平均意义下收敛到随机微分方程的唯一的解.

以上讨论的是取值于实直线R^1的随机微分方程. 对于多变量情形，即取值于R^d的随机微分方程，可以完全类似地得到随机 Taylor 近似及随机 Runge-Kutta 模型.

习题 12

1．求对应于相关函数为$B(t)=g\cdot e^{-b|t|},(g,b>0)$的宽 Markov 平稳 Gauss 过程的条件数学期望

$E(x_{s+t}\mid x_s)$与其相应的条件方差$Var(x_{s+t}\mid x_s)$

2．证明平稳 Gauss-Markov 过程是可逆的．

3．设矩阵 A 为非正定对称（即所有特征值都小于 0），$\{\vec{x}_t: -\infty < t < +\infty\}$ 为 d 维平稳 Gauss 过程，且满足

$$E\vec{x}_t = 0, E(\vec{x}_s \vec{x}_t^T) = e^{A(t-s)} (= I + A(t-s) + \cdots + \frac{A^n (t-s)^n}{n!} + \cdots).$$

证明 $\{\vec{x}_t: 0 \le t < +\infty\}$ 的转移密度 $p(t, \vec{x}, \vec{y})$ 为 $N(e^{At}\vec{x}, I - e^{2At})$（把 $\vec{y}$ 看成分布密度的变量）． 因而有

$$\lim_{t\to\infty} p(t, \vec{x}, \vec{y}) = N(0, I).$$

4. 设正态过程 x_t 的数学期望为 0，协方差函数 $Cov(x_s, x_t) = u(s\wedge t)v(s \vee t)$．证明它是 Markov 过程，又问它是否时齐 ?

5. 若 x_n 是鞅列，则 $x_n \vee c$ 是下鞅列，$x_n \wedge c$ 是上鞅列． 又若 $Ex_n^+ < +\infty$，则 x_n^+ 也是下鞅列．对应地，又若 $Ex_n^- > -\infty$，则 x_n^- 也是上鞅列．

6. 设 x_n 是鞅列，其方差有限，且对于任意 n 有 $x_n = \sum_{k=1}^n X_k$，则序列 $\{X_k\}$ 中的随机变量两两不相关．

7. 在博采输光问题中，如果 $p < q$，求输光概率 p_a, q_b 以及博采至输光为止的平均时间．

8. 设 $\{X_n, n \ge 1\}$ 独立同分布，$X_n \sim \begin{pmatrix} -1 & 1 \\ q & p \end{pmatrix}$，$p > q$．$x_n$ 是简单随机徘徊:$x_n = \sum_{k=1}^n X_k$，

$$h_n = x_n^2 - n[1-(p-q)^2], V_n = \left(\frac{q}{p}\right)^{x_n}, Z_n = x_n - n(p-q).$$

(1) 证明 x_n 是 (X_n) 下鞅列，而 h_n, V_n, Z_n 都是 (x_n) 鞅列．

(2) 求 h_m 与 h_n 的相关系数，

9. 若 $X_n \sim N(m, s^2), x_n = e^{\frac{m}{s^2}\sum_{k=1}^{n} X_k - \frac{nm^2}{2s^2}}$. 证明$x_n$是$(X_n)$鞅列.

10. 假定随机序列$\{x_n, n \geq 0\}$均有数学期望，且满足

$$E(x_{n+1} | x_n, \cdots, x_0] = a x_n + b x_{n-1}, (a, b > 0, a + b = 1).$$

能否选取c，使$h_n = cx_n + x_{n-1}, (n \geq 1, h_0 = x_0)$为$(x_n)$鞅列 ？

11. 设$x_0 \sim U[0,1]$，又在x_n已知的条件下，$x_{n+1} \sim U[1-x_n, 1]$. 证明如下定义的h_n是(x_n)鞅列:

$$\begin{cases} h_0 = x_0 \\ h_n = 2^n \prod_{k=1}^{n} \frac{1-x_k}{x_{k-1}} \end{cases}$$

12. 设$\{x_n, n \geq 0\}$是状态空间$S = \{0,1,2,\cdots,N\}$的时齐 Markov 链.

(1) 若转移概率为$p_{ij} = C_N^j (\frac{i}{N})^j (1-\frac{i}{N})^{N-j}$，那么 $h_n = \frac{x_n(N-x_n)}{(1-\frac{1}{N})^n}$是$(x_n)$鞅列.

(2) 若转移概率为$p_{ij} = \frac{C_{2i}^j C_{2N}^{N-j}}{C_{2N}^N}$，那么存在 l 使 $V_n = \frac{x_n(N-x_n)}{l^n}$ 是 (x_n)鞅列.

13. 设N_t为强度为 l 的 Poisson 过程. 证明$x_t = N_t - l \cdot t$，$h_t = x_t^2 - l \cdot t$，$V_t = e^{-ax_t + l(1-e^{-a})}$都是$(N_t)$鞅.

14. .求 积分 Brown 运动$x_t = \int_o^t B_s ds$的协方差函数 $Cov(x_s, x_t)$ $(0 < s \leq t)$，并证明. (B_t, x_t)是二维 Markov 过程.

15. 设x_t, h_t是鞅， 则$x_t + h_t$是鞅， $\min(x_t, h)_t$是下鞅.

16. 若x_t是 Markov 过程，f为严格单调函数，则$V_t \overset{\Delta}{=} f(x_t)$也是 Markov 过程.

17. .把$e^{aB_t + bt}$写成 Ito 过程的形式.

18 设 $S_n = X_1 + \cdots + X_n$ 是简单随机徘徊: $X_n \sim \begin{pmatrix} -1 & 1 \\ q & p \end{pmatrix} (q = 1 - p)$ 且独立同分布. 证明

$x_n = (pz^{2q} + qz^{-2p})^{-n} z^{S_n - n(p-q)}$ 是鞅列.

19. 利用选样定理证明 Wald 公式 (见第 1 章).

20. 设 B_t 为 Brown 运动, $f(x)$ 为二次连续可微的增函数. 证明

$$x_t = e^{af(B_t) - \frac{1}{2}\int_0^t [a^2 f'(B_s)^2 + af''(B_s)]ds}$$ 是 (B_t) 鞅.

21. 设 $\{l_i\}$ 为有界序列, x_t 是参数为 l_i 的纯生过程. 证明 $h_t = x_t - \int_0^t l_{x_s} ds$, $V_t = e^{ax_t - (1-e^a)\int_0^t l_{x_s} ds}$

都是 (x_t) 鞅.

22 证明 Brown 桥 $X_t = B_t - tB_1 \quad (0 \le t \le 1)$ 是 Markov 过程.

23. 设 $f(t)$ 为平方可积的实值函数, 求证 $X_t = \int_0^t f(s)dB_s$ 为 Gauss 过程, 并计算 $Cov(X_s, X_t)$. 进一步证明 (X_t, Y_t) 是二维 Gauss 过程, 其中 $Y_t = \int_0^t g(t)X_t dt$. 又问 $Cov(X_s, Y_t), Cov(Y_s, Y_t)$ 各是什么 ?

24. 设 $dX_t = (aX_t + b\sqrt{X_t})dt + cX_t dB_t$, $M_t = e^{-cB_t + \frac{1}{2}c^2 t}$,

(1) 求证

$$d(M_t X_t) = M_t(aX_t + b\sqrt{X_t})dt.$$

(2) 令 $Y_t = M_t X_t$, 并证明它满足常微分方程

$$\frac{dY_t}{dt} = aY_t + be^{-\frac{1}{2}cB_t + \frac{1}{4}c^2 t}\sqrt{Y_t}.$$

(3) 求证 $Z_t = \sqrt{Y_t}$ 满足 $\dfrac{dZ_t}{dt} = \dfrac{a}{2}Z_t + \dfrac{b}{2}e^{-\frac{1}{2}cB_t + \frac{1}{4}c^2 t}Z_t$.

(4) 通过求解 Z_t, 证明: $X_t = e^{cB_t - \frac{c^2}{2}t + at}\left(\frac{b}{2}\int_0^t e^{-\frac{c}{2}B_s + (\frac{c^2}{4} - \frac{a}{2})s} ds + \sqrt{X_0}\right)^2$.

25. 设 $f(t, x)$ 为对 x 二阶光滑(二阶导数连续)且对 t 一阶光滑(一阶导数连续)的实函数

$h_t = e^{\int_0^t x_s ds} f(t, x_t)$．求 dh_t ．

26. 求证如下的随机利率的方程 $dr_t = s \cdot dB_t + (a - br_t)dt$ 有唯一解

$$r_t = r_0 e^{-at} + b(1 - e^{-at}) + s \cdot e^{-at} \int_0^t e^{as} dB_s .$$

再求它的期望函数 Er_t 与协方差函数 $Cov(r_s, r_t)$.

27. d 维 Brown 运动 $(B_t^{(1)}, \cdots, B_t^{(d)})$ 的向径过程 $R_t = \sqrt{(B_t^{(1)})^2 + \cdots + (B_t^{(d)})^2}$ 称为 Bessel 过程. 请利用多维 Ito 公式, 导出它所满足的随机微分方程.

28. 由定义求 $\int_0^t s dB_s$, $\int_0^t B_s^2 dB_s$.

29. 求由原点出发的二维 Brown 运动在时刻 t 落在圆盘 $x^2 + y^2 \le r^2$ 中的概率.

30. 判别下列是否为鞅 (需要证明) ：$t^2 B_t - 2\int_0^t sB_s ds$, $B_t^3 - 3tB_t$

31 设 $(B_t^{(1)}, B_t^{(2)})$ 是二维 Brown 运动. 判别下列是否为鞅 （需要证明）：$B_t^{(1)} B_t^{(2)}$ ，$\ln([B_t^{(1)}]^2 + [B_t^{(2)}]^2)$.

32 归纳法证明 ；$\frac{n!}{t^{\frac{n}{2}}} \int \cdots \int_{0 \le u_1 \le \cdots \le u_n} dB_{u_1} \cdots dB_{u_n} = h_n(\frac{B_t}{\sqrt{t}})$，其中 $h_n(x)$ 是一个多项式， 称为 n 阶 Hermite 多项式.

33. 将下列 Stratonovich 方程化为 Ito 方程：$dX_t = bX_t dt + aX_t \circ dB_t$，

$dX_t = \sin X_t \cos X_t dt + \{t^2 + \cos X_t) \circ dB_t$。

34. 将下列 Ito 方程化为 Stratonovich 方程：$dX_t = bX_t dt + aX_t dB_t$， $dX_t = 2e^{-X_t} dt + X_t^2 dB_t$ 。

35. 求 dX_t， 其中 $X_t = B_t^n$， $X_t = 2 + t + e^{B_t}$ ， $X_t = [B_t^{(1)}]^2 + [B_t^{(2)}]^2$, $(B_t^{(1)}, B_t^{(2)})$ 是二维 Brown 运动.

36. 设 B_t 是三维 Brown 运动。证明 $X_t = \frac{1}{|x + B_t|}$ 是鞅.

37. 记 $g_k(t) = E|B_t^k|$。证明 $g_k(t) = \frac{1}{2}k(k-1)\int_0^t g_{k-1}(s) ds \quad (k \ge 2)$.

38. 设 $X_t = e^{ct+\sum_{k=1}^{3} b_k B_t^{(k)}}$ ，其中$(B_t^{(1)}, B_t^{(2)}, B_t^{(3)})$是三维 Brown 运动，求 dX_t .

39. 设 $g_e(x) = |x| I_{\{|x|\geq e\}} + \frac{1}{2}(e + \frac{x^2}{e}) I_{\{|x|<e\}}$. 证明 Tanaka 公式

$$g_e(B_t) = g_e(B_0) + \int_0^t g_e'(B_s) dB_s + \frac{1}{2e} |\{s : 0 \leq s \leq t, -e < B_s < e)|,$$

其中 $|\{s : 0 \leq s \leq t, -e < B_s < e)|$ 表示集合 $\{s : 0 \leq s \leq t, -e < B_s < e)$ 的测度.

40. 求解 $dX_t = (AX_t + C)dt + KdB_t$.

41. 设 $g(x) = (\cos x, \sin x)$ ，求$(X_t, Y_t) = g(B_t)$ (称为圆上的 Brown 运动)满足的随机微分方程.

42. 证明$\sin B_t$是方程$dX_t = -\frac{1}{2} X_t dt + \sqrt{1 - X_t^2} dB_t \quad (B_0 = x \in (-\frac{p}{2}, \frac{p}{2}))$ 的解.

43. 求下列方程的解 （1）$d\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} 1 \\ Y_t \end{pmatrix} dt + \begin{pmatrix} 0 \\ e^{X_t} \end{pmatrix} dB_t$ （提示 $\begin{pmatrix} t \\ e^t B_t \end{pmatrix}$）.

（2）$d\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \frac{1}{2}\begin{pmatrix} X_t \\ Y_t \end{pmatrix} dt + \begin{pmatrix} Y_t \\ X_t \end{pmatrix} dB_t$ （提示 $\begin{pmatrix} chB_t \\ shB_t \end{pmatrix}$）.

（3）$d\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \frac{1}{2}\begin{pmatrix} X_t \\ Y_t \end{pmatrix} dt + \frac{b}{a}\begin{pmatrix} -Y_t \\ X_t \end{pmatrix} dB_t$ （提示 $\begin{pmatrix} a\cos B_t \\ b\sin B_t \end{pmatrix}$）.

（4）$d\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} dt + \begin{pmatrix} dB_t^{(1)} \\ X_t dB_t^{(2)} \end{pmatrix}$.

44. 对于下列方程求解，并求其转移密度满足的向后方程与向前方程

（1）$dX_t = X_t dt + dB_t$. （2）$dX_t = -X_t dt + e^{-t} dB_t$.

（3）$dX_t = rX_t dt + \sum_{k-=1}^{m} s_k dB_t^{(k)}$. （4）$dX_t = \ln(1 + X_t^2)dt + X_t I_{(0,\infty)}(X_t) dB_t$.

龚光鲁，钱敏平著　应用随机过程教程 – 与在算法和智能计算中的应用
清华大学出版社，2003

第 13 章 金融证券未定权益的定价

1　Black-Scholes 模型的欧式未定权益的定价

1．1　术语与基本假定

概念１３．１　（可行市场）　研究金融市场有一个基本的假定，就是无套利原则，也称套利原则，这个原则就是假定正常运行的市场没有套利机会．(套利的粗略含义是，在开始时无资本，经过资本的市场运作后，变成有非负的（随机）资金，而且有正资金的概率为正)．因为在出现套利机会时，大量的投机者就会涌向市场进行套利，于是经过一个相对短的时期的″混乱″后，市场就会重返″正常″，即回复到无套利状态．在金融衍生证券的定价理论中并不讨论这段短混乱时期，因此，在研究中，普遍地设置无套利假定，这样的市场也称为可行市场．

概念１３．２（套期）　粗略地说，以持有某些有价证券组合来抵销某种金融衍生证券所带来的风险，称为套期，这种套期事实上是完全套期．如果只抵销了部分风险，则称为部分套期．

定义１３．３（欧式期权，欧式未定权益）　设某种风险金融证券每份在t时刻的价格为S_t，并设它满足以下的 Black-Scholes 模型：

$$dS_t = S_t(m dt + s\, dB_t)\,, \tag{13．1}$$

其中$m, s\ (>0)$分别为证券的收益率与波动率．假定当前的银行利率为r，而且不随时间变化．以这种证券为标的变量(Underlying Variable)的欧式看涨期权（European call option)，是指在$t=0$时甲方(一般为证券公司)与乙方的一个合约，按此合约规定乙方有一个权利，能在时刻T以价格K（它称为敲定价格，striking price）从甲方买进一批(一般为 100 份)这种证券．如果时间T时的市场价格S_T低于K，乙方可以不买，而只要时间T时的市场价格S_T高于K，乙方就得利．综合起来，乙方在时刻T净得随机收益为$X_T=(S_T-K)^+$．这种合约（它的数学表示就是$X_T=(S_T-K)^+$）称为期权．又因为乙方只能在最终时刻T作出选择，称为欧式期权．此外，乙方希望S_T尽量大，以便有更多的获利．也就是有选择权的乙方盼望股票上涨，所以称为看涨期权，或者买权．由于这个合约能给乙方带来X_T的随机收益，就需要乙方在$t=0$时刻用钱从甲方购买．这个合约在$t=0$时刻的价格，称为它的贴水或保证金(premium)．问题是如何确定这个合约在时刻$t<T$的价格(包括贴水)．

另一种相反的情况是，如果$t=0$时甲方(一般为证券公司)卖给乙方如下的合约，此合约规定乙方有一个权利,即能在时刻T以价格为K卖给甲方一批(一般也为 100 份)这种证券．如果时刻T时的市场价格S_T高于K，乙方可以不卖．只要时间T时的市场价格S_T低于K，

卖方就得利．综合起来，乙方在时刻T净得为随机收益$X_T=(K-S_T)^+$．这也是一种欧式期权，此时乙方盼望S_T尽量小，以便有更多的获利．也就是，乙方盼望股票下跌，所以称为看跌期权，或者卖权 (put option)．同样由于这个合约也能给乙方带来X_T的收益，也就需要乙方在$t=0$时刻用钱从甲方购买．这个合约在$t=0$时刻的价格，也称为它的贴水或保证金．

比看涨期权与看跌期权更为一般的欧式期权是：甲方卖给乙方一个由证券组合组成的一个合约，此合约能在T时刻给乙方带来随机收益$f(S_T)$(称为欧式未定权益)，同样要给出这个合约在时刻$t<T$的价格(贴水).

基本假定　为了讨论简单，通常假定市场是无摩擦的，即无税收，无交易费，可以卖空，银行的存贷利率是一样的．又假定此未定权益的持有人是小投资者，并且是自融资的，即在整个过程中，持有人既没有添入资金，也没有抽走资金．

再则，存银行的钱是无风险的．由银行利率为 r (在时变情形为$r(t)$，这在数学处理上并未增加任何困难)，时刻$t=0$的存款S_0^0元到时刻t的价值应为

$$S_t^0=S_0^0e^{rt}，\text{即} dS_t^0=rS_t^0dt. \tag{13. 2}$$

无套利下的看涨与看跌平权关系

定义１３．４（远期合约）　未定权益为S_T的欧式权益，称为在时刻T成熟的远期合约．显见远期合约在时刻$t(<T)$的价格就应该为证券的即时价格S_t．

命题１３．５（平权关系）将把看涨期权，看跌期权，远期合约在时刻$t(<T)$的价格分别记为C_t,P_t,F_t．于是在无套利假定下有

$$C_t-P_t=F_t-Ke^{-rt}. \tag{13. 3}$$

这个关系式称为欧式看涨-看跌期权的平权关系(Call-Put parity)．有了这个关系，欧式看涨期权与看跌期权中只要知道一个的价格，就立刻可以得到另一个的价格．

(平权关系的得到，是基于未定权益有如下的等式:

$$(S_T-K)^+-(K-S_T)^+=S_T-K.$$

这说明买进一张(在金融中称为多头一张)在T到期的执行价格为K的看涨期权与卖出一张(在金融中称为空头一张)相应的看跌期权，就相当于买进一张远期合约与卖出一张在时刻T到期的额度为K的银行存款．

1．2　定价的套期方法

１．Black-Scholes 偏微分方程的推导

满足 Black-Scholes 模型（１３．１）的风险标的资产S_t是随机微分方程的解，因此是

Markov 过程. 从而, 欧式未定权益 $f(S_T)$ 在时刻 $t<T$ 的价格只依赖 S_t, 于是可以将此价格随机过程记为 $V(t,S_t)$. 下面我们将按 Black-Scholes 在 1973 年的经典论文中的套期思想, 来推导价格函数 $V_t \overset{\Delta}{=} V(t,x)$ 满足的微分方程. 首先, 由 Ito 公式有

$$dV(t,S_t)=[\frac{\partial V}{\partial x}\sigma\cdot S_t dB_t+(\frac{\partial V}{\partial t}+\mu\cdot S_t\frac{\partial V}{\partial x}+\frac{1}{2}\sigma^2 S_t^2\frac{\partial^2 V}{\partial x^2})dt]_{x=S_t}.$$

假定在时刻 t 甲方为了保护卖出未定权益的风险, 虚拟地待定购进 Δ 份标的证券. 那么, 将卖出的未定权益合约得到的价格 V_t 与购进的标的证券化费的 $\Delta\cdot S_t$ 合起来考虑, 甲方共计得资产

$$N_t \overset{\Delta}{=} V_t-\Delta\cdot S_t.$$

由自融资原则, 到了时刻 $t+dt$,这个获利就变为 $N_{t+dt} \overset{\Delta}{=} V_{t+dt}-\Delta\cdot S_{t+dt}$. 即

$$dN_t=dV-\Delta\cdot dS_t=[(\frac{\partial V}{\partial x}-\Delta)\sigma\cdot S_t dB_t+(\frac{\partial V}{\partial t}+\mu\cdot S_t(\frac{\partial V}{\partial x}-\Delta)+\frac{1}{2}\sigma^2 S_t^2\frac{\partial^2 V}{\partial x^2})dt]_{x=S_t}.$$

我们只要取

$$\Delta=\frac{\partial V}{\partial x}(t,S_t), \qquad (13.\ 4)$$

便得到

$$dN_t=(\frac{\partial V}{\partial t}+\frac{1}{2}\sigma^2 S_t^2\frac{\partial^2 V}{\partial x^2})_{x=S_t}dt. \qquad (13.\ 5)$$

再由无套利假定可知, N_t 必须是无风险资产, 即

$$dN_t=rN_t dt. \qquad (13.\ 6)$$

(事实上, 如果 $dN_t>rN_t dt$, 则甲可以在时刻 t 从银行借贷并投资于上述组合 N_t, 在 $t+dt$ 时刻得 dN_t 后立刻偿还银行 $rN_t dt$, 从而净得 $dN_t-rN_t dt$. 在另一种情形，如果 $dN_t<rN_t dt$, 则甲可以在时刻 t 卖空上述组合 N_t, 将钱存入银行, 到 $t+dt$ 时刻得 $rN_t dt$, 还去购回卖空的组合, 由此也净得 $rN_t dt-dN_t$. 所以, 这两种情形都发生了套利, 与无套利假定矛盾. 可见只能有 $dN_t=rN_t dt$). 把 (13. 5)式中的 N_t 的表达式代入(13. 6)式, 就得到 $V(t,x)$ 应满足的 Black-Scholes 偏微分方程:

$$\frac{\partial V}{\partial t}+\frac{1}{2}s^{2}S_t^2\frac{\partial^2 V}{\partial x^2}+rx\frac{\partial V}{\partial x}-rV=0, \qquad (13.7)$$

外加终端条件

$$V(T,x)=f(x)\text{，（因为}V(T,S_T)=f(S_T)\text{）.} \qquad (13.8)$$

只要求得此方程的解 $V(t,x)$，**就得到了未定权益** $f(S_T)$ **在时刻** $t<T$ **的价格** $V(t,S_t)$.

而在时刻 $t=0$ **贴水为** $V(0,S_0)$.

２． Black-Scholes **微分方程的求解**

先令 $t'=T-t$ **将终值问题化成初值问题，再令** $x'=\log x$, $\tilde{V}(t',x')=V(t,x)$ **， 并利用**

$$\tilde{V}_{x'}=xV_x, \tilde{V}_{x'x'}=(xV_x)_{x'}=x(xV)_x=x^2V_{xx}+xV_x, \ \tilde{V}_{t'}=-V_t,$$

把方程（１３．７）化为

$$-\tilde{V}_{t'}+\frac{1}{2}s^{2}\tilde{V}_{x'x'}+(r-\frac{s^2}{2})\tilde{V}_{x'}-r\tilde{V}=0, \qquad (13.9)$$

$$\tilde{V}(0,x')=(V(T,x)=f(x)=)f(e^{x'})\text{.} \qquad (13.10)$$

此方程的系数不依赖 x' **，是常系数偏微分方程．然后作变换**

$$\tilde{V}(t',x')=e^{ax'+b\cdot t'}U(t',x'). \qquad (13.11)$$

只要在此变换中选取合适的常数 a,b **，就可以消灭方程中含** U **和含** $U_{x'}$ **的项,即**

（１）置 $U_{x'}$ **的系数为** 0，**得到**

$$a=-\frac{r-\frac{s^2}{2}}{s^2}=\frac{1}{2}-\frac{r}{s^2}, \qquad (13.12)$$

（２）再置 U **的系数为** 0，**得到**

$$b=-(\frac{1}{2}s^2a^2+r). \qquad (13.13)$$

这时（１３．９)，（１３．１０）就简化为传热方程的初值问题

$$U_{t'}=\frac{s^2}{2}U_{x'x'} \qquad (13.14)$$

$$U(0,x')=e^{-ax'}f(e^{x'}). \qquad (13.15)$$

由此可以用初等偏微分方程中的经典方法，即用 Gauss **核**（Brown **运动的转移密度函数也称为** Gauss **核**）**的积分给出其解** $U(t',x')$

$$U(t',x')=\frac{1}{\sqrt{2p\cdot t's}}\int U(0,y)e^{-\frac{(y-x')^2}{2t's^2}}dy$$

$$=\frac{1}{\sqrt{2p\cdot t's}}\int e^{-ay}f(e^y)e^{-\frac{(y-x')^2}{2t's^2}}dy\text{.}$$

再作变换 $y-x'=y'$，便得

$$U(t',x')=\frac{1}{\sqrt{2p\cdot t's}}\int f(e^{y'+x'})e^{-\frac{y'^2}{2t's^2}}e^{-a(x'+y')}dy'=\frac{e^{-ax'}}{\sqrt{2p\cdot t's}}\int f(xe^{y'})e^{-\frac{y'^2+2t's^2ay'}{2t's^2}}dy'.$$

注意积分号内指数项的分子为 $(y'+t's^2a)^2-t'^2s^4a^2$，作变换 $\frac{y'+t's^2a}{s\sqrt{t'}}=z$，就得

$$U(t',x')=\frac{e^{-ax'}}{\sqrt{2p}}\int f(xe^{s\sqrt{t'}z-t's^2a})e^{-\frac{z^2}{2}}e^{\frac{t's^2a^2}{2}}dz$$

$$=\frac{e^{-ax'}e^{-(b\cdot t'+rt')}}{\sqrt{2p}}\int f(xe^{s\sqrt{T-t}z-(T-t)s^2a})e^{-\frac{z^2}{2}}dz\text{.}$$

最后用 $V(t,x)=\tilde{V}(t',x')=e^{ax'+b\cdot t'}U(t',x')$，得到 $V(t,x)$ 的明显表达式如下面定理所述.

定理１３．６

$$V(t,x)=\frac{e^{-r(T-t)}}{\sqrt{2p}}\int f(xe^{s\sqrt{T-t}z-(T-t)s^2(r-\frac{s^2}{2})})e^{-\frac{z^2}{2}}dz\text{.}\qquad(13.\ 16)$$

于是欧式未定权益 $f(S_N)$ 在时刻 $t(<T)$ 的价格为 $V(t,S_t)$，而在合约开始时刻的贴水为

$$V(0,S_0)=\frac{e^{-rT}}{\sqrt{2p}}\int f(S_0e^{s\sqrt{T}z-Ts^2(r-\frac{s^2}{2})})e^{-\frac{z^2}{2}}dz\text{.}\qquad(13.\ 17)$$

而用以套期标的证券的数量，则由(13. 4)式给出．

定义１３．７　（Black-Scholes 中性模型） 从(13. 16)可以看出，欧式未定权益的定价不依赖风险证券的收益率 m．而代替它的则是银行利率 r．这就启示我们，在利用 Black-Scholes 模型求未定权益的定价时，风险证券的价格模型应该改用

$$dS_t=S_t(rdt+dB_t).\qquad(13.\ 18)$$

这样的模型称为 Black-Scholes 风险中性模型．由于它与收益率为 m 时的模型不一样，我们把对应于这个风险中性模型所取的概率记为 P^* (相应的期望记为 E^*)，称为风险中性概率．

在金融中，$\frac{m-r}{s}$ 称为风险的市场价格．

１．３　风险中性概率方法

设在风险中性模型下，风险证券在 t 时刻的价格记为 S_t．它对于银行利率 r 的折现价

记为 $\tilde{S}_t \overset{\Delta}{=} e^{-rt}S_t$ (在初始时刻的等价价格，相当于将一切资产都折合到初始时刻考虑). 用 Ito 公式得

$$d\tilde{S}_t = d(e^{-rt}S_t) = -re^{-rt}S_t dt + e^{-rt}S_t(rdt + s\,dB_t) = s\,\tilde{S}_t\,dB_t. \qquad (13.\ 19)$$

由此可见对风险中性概率 P^* 而言，风险证券的折现价 $\tilde{S}_t$ 是随机微分方程（１３．１９）的解，由例１２．５７，此方程的唯一解是

$$\tilde{S}_t = \tilde{S}_0\, e^{-\frac{s^2}{2}t + sB_t}, \qquad (13.\ 20)$$

这正是 B_t 的一个指数鞅． 因为鞅在时间进行中体现了公平的发展， 在时刻 T 的未定权益 $f(S_T)$ 在初始时刻的折现价为 $\tilde{f} \overset{\Delta}{=} e^{-rT} f(e^{rT}\tilde{S}_T)$，并且鞅性质体现为，时刻 t 的价格在初始时刻的折现价 $\tilde{V}_t \overset{\Delta}{=} e^{-rt}V_t$ 应该是 $E^*(\tilde{f} \mid B_s, s \le t)$，即

$$\tilde{V}_t = E^*(\tilde{f} \mid B_s, s \le t). \qquad (13.\ 21)$$

于是

$$V_t = e^{rt}E^*(e^{-rT} f(e^{rT}\tilde{S}_T) \mid B_s, s \le t)$$

$$= e^{-r(T-t)}E^*[f(S_t e^{s(B_T - B_t) + (r - \frac{s^2}{2})(T-t)}) \mid B_s, s \le t]. \qquad (13.\ 22)$$

再利用条件期望的公式，并令 $h \overset{\Delta}{=} \dfrac{B_T - B_t}{\sqrt{T-t}}$，则对应于概率 P^* 有 $h \sim N(0,1)$．故而

$$V_t = [e^{-r(T-t)}E^*(f(xe^{s(B_T - B_t)}))]_{x = S_t \exp[(r - \frac{s^2}{2})(T-t)]}$$

$$= [e^{-r(T-t)}E^*(f(xe^{s\sqrt{T-t}\cdot h}))]_{x = S_t \exp[(r - \frac{s^2}{2})(T-t)]}$$

$$= e^{-r(T-t)}\frac{1}{\sqrt{2p}}\int f(S_t e^{(r - \frac{s^2}{2})(T-t) + s\sqrt{T-t}\cdot u})e^{-\frac{u^2}{2}}du \overset{\text{记为}}{=} V(t, S_t),$$

其中

$$V(t,x) = e^{-r(T-t)}\frac{1}{\sqrt{2p}}\int f(xe^{(r - \frac{s^2}{2})(T-t) + s\sqrt{T-t}\cdot u})e^{-\frac{u^2}{2}}du.$$

它正好是带终端条件的 Black-Scholes 偏微分方程的解(13. 16)．

推论１３．７ 对于远期合约 $f(x) = x$，由 Black-Scholes 公式得到它在时刻 $t(t < T)$ 的价格为 $F(t, S_t) = S_t$，恰好就是标的风险证券的市价.

推论１３．８ 对于欧式看涨期权（对应于 $f(x) = (x - K)^+$），它在时刻 $t(t < T)$ 的

价格的公式可以简化为

$$F(t,S_t)=x\Phi(d_1(x))-Ke^{-r(T-t)}\Phi(d_2(x)),$$

其中

$$d_1(x)=\frac{\log(\frac{x}{K})+(r+\frac{s^2}{2})(T-t)}{s\sqrt{T-t}},\quad d_2(x)=\frac{\log(\frac{x}{K})+(r-\frac{s^2}{2})(T-t)}{s\sqrt{T-t}},$$

$\Phi(x)=\frac{1}{\sqrt{2p}}\int_{-\infty}^{x}e^{-\frac{u^2}{2}}du$ (在通常的金融文献中记为 $N(x)$) 是为标准正态分布的分布函数.

此时应该套期的风险证券的数量为

$$\Delta_t=\frac{\partial F}{\partial x}|_{x=S_t}=[\Phi(d_1(x))+xe^{-\frac{d_1(x)^2}{2}}d_1'(x)-Ke^{-r(T-t)}e^{-\frac{d_2(x)^2}{2}}d_2'(x)]_{x=S_t}=\Phi(d_1(S_t)).$$

推论１３，９ 由平权关系便得到欧式看跌期权（对应于 $f(x)=(K-x)^+$）在时刻 $t(t<T)$ 的价格

$$F(t,S_t)=Ke^{-r(T-t)}\Phi(-d_2(x))-x\Phi(-d_1(x)).$$

此时应该套期的风险证券的数量为

$$\Delta_t=\frac{\partial F}{\partial x}|_{x=S_t}=\cdots=-\Phi(-d_1(S_t)).$$

[注] 如果不用风险中性的 Black-Scholes 模型，而是用带收益率 m 的 Black-Scholes 模型，那么类似地用 Ito 公式，可得证券的折现价格 $\tilde{S}_t$ 满足的方程为:

$$d\tilde{S}_t=\tilde{S}_t s(dB_t+\frac{m-r}{s}dt).$$

作 Girsanov 变换 $B^*=B_t+\frac{m-r}{s}t$，则对于 F 中的任意事件 A，只要它的信息完全可由 $\{B_s:s\le t)$ 确定，就定义它的一个新概率 P^* 为

$$P^*(A)=E(I_Ae^{-\frac{m-r}{s}B_t-\frac{1}{2}\left(\frac{m-r}{s}\right)^2t}).$$

于是由 Girsanov 定理（定理１２．６９），在新概率 P^* 下，B^* 是一个 Brown 运动．从而重新得到 $d\tilde{S}_t=\tilde{S}_t s\cdot dB_t^*$. 这正好是风险中性的 Black-Scholes 模型．而概率 P^* 就是风险中性概率．这是从另一个角度得到同样的结果．

*1．.4 币值单位与随机折现因子方法

定义１３．１０ 设某个风险证券在时刻 t 的价格为 S_t. 又若存在另一个正值随机过程 M_t 使得

$\tilde{S}_t \stackrel{\Delta}{=} \frac{S_t}{M_t}$ 是一个鞅，则称 M_t 为证券 S_t 的币值单位（ 更一般一些，在理论上只要 $\tilde{S}_t$ 是一个局部鞅，也

就是存在一个停时（$\tilde{S}_t$）序列 $t_n \uparrow \infty$，使对于固定的 n，$\{\tilde{S}_{t \wedge t_n} : t \geq 0\}$ 是鞅． 这里， 鞅和局部鞅都代表在统计平均意义下不随时间增值的资金流）．

从直观看，资本的价值是随着时间增值的，币值单位 M_t 就体现了该证券的价格的时间增值，即在 $t=0$ 时的 1 元钱，经过单个证券 S_t 的市场因素的作用，在时刻 t 的实际价值相当于 M_t 元．又因为市场是随机的，所以币值单位也是随机的量(随机过程)，这就体现了一元钱的时间价值．

定义１３．１１

$$N_t \stackrel{\Delta}{=} \frac{1}{M_t} \tag{13. 23}$$

称为随机折现因子，因为 N_t 是 1 除以币值单位，所以它就相当于把随机因素考虑进去以后的 ”随机折现因子”．

对于 Black-Scholes 模型的币值单位， 我们有下面的命题．

命题１３．１２ 假定证券在时刻 t 的价格 S_t 满足 Black-Scholes 模型

$$dS_t = S_t(m dt + s dB_t). \tag{13. 24}$$

那么

$$M_t = e^{rt} L_t, \tag{13. 25}$$

就是 S_t 的币值单位，其中 r 是银行利率，而

$$L_t \stackrel{\Delta}{=} e^{\frac{m-r}{s}B_t + \frac{1}{2}\frac{(m-r)^2}{s^2}t} \stackrel{\Delta}{=} e^{G(t,B_t)}. \tag{13. 26}$$

证明 对于 L_t 用 Ito 公式得到

$$dL_t = e^{G(t,B_t)} dG(t,B_t) + \frac{1}{2} e^{G(t,B_t)} (dG(t,B_t))^2$$
$$= e^{G(t,B_t)} [\frac{m-r}{s} dB + (\frac{m-r}{s})^2 dt].$$

于是 M_t 及 $N_t \stackrel{\Delta}{=} \frac{1}{M_t}$ 满足的随机微分就分别为

$$dM_t = re^{rt} dt L_t + e^{rt} dL_t = rM_t dt + M_t [\frac{m-r}{s} dB_t + (\frac{m-r}{s})^2 dt]$$
$$= M_t([r + (\frac{m-r}{s})^2] dt + \frac{m-r}{s} dB_t).$$

$$dN_t = -\frac{1}{M_t^2} dM_t + \frac{1}{M_t^3} (dM_t)^2$$

$$= -\frac{1}{M_t}([r + (\frac{m-r}{s})^2]dt + \frac{m-r}{s} dB_t) + \frac{1}{M_t^3} M_t^2 (\frac{m-r}{s})^2 dt$$

$$= -N_t (rdt + \frac{m-r}{s} dB_t).$$

即 N_t 所满足的随机微分方程为:

$$dN_t = N_t(-rdt - \frac{m-r}{s} dB_t). \qquad (11.\ 27)$$

再用 Ito 公式得到

$$d(S_t N_t) = S_t dN_t + N_t dS_t + dS_t dN_t$$

$$= S_t N_t (-rdt - \frac{m-r}{s} dB_t) + N_t S_t (mdt + s\,dB_t) + S_t s N_t (-\frac{m-r}{s}) dt$$

$$= S_t N_t (s - \frac{m-r}{s}) dB_t.$$

由例１２．５７可知，$\frac{S_t}{M_t} = S_t N_t$ 是一个指数鞅．命题得证．

既然 $\tilde{S}_t \overset{\Delta}{=} \frac{S_t}{M_t}$ 随时间的发展体现了一个 ”公平博弈”， 由此可得到，一般的欧式未定权益 X_T 在 $t(<T)$ 的合理定价 V_t，即对于折现价格 $\tilde{V}_t \overset{\Delta}{=} \frac{V_t}{M_t}$， $\tilde{X}_T \overset{\Delta}{=} \frac{X_T}{M_T}$，应该有

$$\tilde{V}_t = E(\tilde{X}_T \mid B_s : s \le t).$$

从而得到一般的定价公式

$$V_t = M_t E(\frac{X_T}{M_T} \mid B_s, s \le t) = e^{rt} L_t E(X_T e^{-rT} \frac{1}{L_T} \mid B_s, s \le t). \qquad (13.\ 28)$$

1．5　倒向随机微分方程方法

设风险证券的价格过程为 S_t．将由它衍生的欧式未定权益 $f(S_T)$ 在 t 时刻的价格记为 x_t．利用套期的想法，在 t 时刻“ 虚拟 ”地待定卖出 n_t 张标的风险证券(这里 n_t 可以负，此时表示买进)，使余下的部分不再有风险．在“ 卖出 ” 了 n_t 张风险证券后的投资组合为 $x_t - n_t S_t$．由自融资假定，在时刻 $t+dt$ 这个投资组合的价值变为 $x_{t+dt} - n_t S_{t+dt}$．又因为要求它是无风险的，所以它应该相当于时刻 t 的资产 $x_t - n_t S_t$ 存进银行中到时刻 $t+dt$ 的资产，即

$$x_{t+dt}-n_tS_{t+dt}=(1+rdt)(x_t-n_tS_t)\,.\qquad(13.29)$$

用 dx_t, dS_t 分别代替 $x_{t+dt}-x_t$，$S_{t+dt}-S_t$ 后，上式成为

$$dx_t-n_tdS_t=r(x_t-n_tS_t)dt\,.\qquad(13.30)$$

再用 $dS_t=S_t(mdt+s\,dB_t)$ 代入，并把总的套期的价值记为

$$V_t=n_tS_t\,.\qquad(13.31)$$

那么，我们就得到下述的倒向随机微分方程：

$$dx_t=(rx_t+(m-r)V_t)dt+s\cdot V_tdB_t\,,\qquad(13.32)$$

$$x_T=f(S_T)\,.\qquad(13.33)$$

注意，这里是一个方程，一个终端条件，两个未知的随机过程 (x_t,V_t). 这是倒向随机微分方程所特有的. 这一点是与普通(正向的)随机微分方程是绝然不同的. 正向的随机微分方程对一个初始条件，在一个未知随机过程时有唯一解，而倒向随机微分方程对一个终端条件，只有在两个未知过程时，才可能存在唯一解 (x_t,V_t). 这里，x_t 是证券在时刻 t 的价格，而 V_t 则是在时刻 t 的套期数量.

在数学上，倒向随机微分方程与经典的情形的不同的是，要求解 (x_t,V_t) 是 (B_t) 可知的. 彭实戈等人的一般理论指出：只要满足 $Ef(S_T)^2<\infty$， 方程(13.32)在(13.33)条件下存在唯一的解 (x_t,V_t). 于是这时 x_0 (它将是常数，不再随机) 是欧式未定权益 $f(S_T)$ 在开始时刻 $t=0$ 时的贴水. 倒向随机微分方程的有效而快速的数值解法，目前是研究的关心点之一.

1．6 时变的 Black-Scholes 模型

设风险证券在 t 时刻的价格 S_t 满足

$$dS_t=S_t(m(t)dt+s(t)dB_t)\,,\qquad(13.34)$$

其中 $m(t), s(t)(>0)$ 是非随机的函数，分别代表证券的时变收益率与波动率. 基于这种模型的证券的欧式未定权益的定价，在用上述 4 种方法中的任意一种后，都可以最后得到

$$V(t,x)=e^{-r(T-t)}\frac{1}{\sqrt{2p}}\int f\left(xe^{\int_t^T(r(s)-\frac{s(s)^2}{2})ds+\frac{1}{\sqrt{T-t}}\int_t^T s(s)ds\cdot z}\right)e^{-\frac{z^2}{2}}dz\,.\qquad(13.35)$$

从而，贴水为

$$V(0,S_0)=e^{-rT}\frac{1}{\sqrt{2p}}\int f\left(S_0e^{\int_0^T(r(s)-\frac{s(s)^2}{2})ds+\frac{1}{\sqrt{T}}\int_0^T s(s)ds\cdot z}\right)e^{-\frac{z^2}{2}}dz\,.$$

2．二叉模型与 Black-Scholes 模型的二叉近似

2．1　二叉模型

二叉模型(Cox-Ross-Rubinstein 模型–相当于离散的 Black-Scholes 模型)与风险中性概率.

二叉模型考虑的是时间离散情形，它比 Black-Scholes 模型更适合于实际拟合与模拟计算．设银行利率为 R (注意离散时间的利率与连续实际的利率之间有一个折合关系)， 即存在银行的资金的增长为

$$S_{n+1}^0 = (1+R)S_n^0 . \qquad (13.36)$$

假定风险证券价格的增长为

$$S_{n+1} = (1+h_n)S_n ,$$

其中 $\{h_n\}$ 是独立同分布的随机变量列，且

$$1+h_n \sim \begin{pmatrix} a & b \\ p & 1-p \end{pmatrix}, (0<a<b, 0<p<1),$$

即

$$\frac{S_{n+1}}{S_n} \sim \begin{pmatrix} a & b \\ p & 1-p \end{pmatrix}, \qquad (13.37)$$

此时，证券在时刻 $t=0$ 的折现价格为

$$\tilde{S}_n \overset{\Delta}{=} \frac{1}{(1+R)^n} S_n . \qquad (13.38)$$

由 Black-Scholes 模型启示，无套利相当于对于 $\frac{S_{n+1}}{S_n}$ 存在一个风险中性的概率分布

$$\begin{pmatrix} a & b \\ p^* & 1-p^* \end{pmatrix}$$

(p^* 称为公平概率，对应于事件类 F 上定义的风险中性概率 $P^*(\cdot)$)， 它使风险证券的折现价格 $\tilde{S}_n$ 关于 P^* 是鞅列，即

$$E^*(\tilde{S}_{n+1} \mid \tilde{S}_n, \cdots, \tilde{S}_0) = \tilde{S}_n .$$

定理１３．１３　当且仅当在 $a<1+R<b$ 时，公平概率 p^* 存在唯一，在条件成立时有

$$p^* = \frac{b-(1+R)}{b-a}。 \qquad (13.39)$$

证明 　由于 $\{\frac{S_{n+1}}{S_n}\}$ 的独立同分布性推出 $\frac{\tilde{S}_{n+1}}{\tilde{S}_n}=\frac{1}{1+R}\frac{S_{n+1}}{S_n}$ 与 $\{\tilde{S}_n,\cdots,\tilde{S}_0\}$ 是相互独立的. 于是

$$\tilde{S}_n=E^*(\tilde{S}_{n+1}\mid\tilde{S}_n,\cdots,\tilde{S}_0)=E^*(\frac{\tilde{S}_{n+1}}{\tilde{S}_n}\tilde{S}_n\mid\tilde{S}_n,\cdots,\tilde{S}_0)=\tilde{S}_nE^*(\frac{\tilde{S}_{n+1}}{\tilde{S}_n}).$$

从而有

$$1=E^*(\frac{\tilde{S}_{n+1}}{\tilde{S}_n})=\frac{a}{1+R}p^*+\frac{b}{1+R}(1-p^*),$$

解出 $p^*=\frac{b-(1+R)}{b-a}$, 所以，为了使 $0<p^*<1$ 必须满足 $a<1+R<b$. 　】

二叉模型下的欧式未定权益的定价

由处理 Black-Scholes 模型的经验，我们可以假定未定权益 $f(S_N)$ 在时刻 $n(<N)$ 的价格为 $V_n\overset{\Delta}{=}V(n,S_n)$. 我们沿用套期方法计算它. 设在时刻 n 卖空此未定权益，并待定买进 Δ_n 张标的证券，并把此投资组合的价值记为 X_n, 则

$$X_n=\Delta_nS_n-V_n.$$

在 S_n 已知的条件下， 要用此投资组合来抵销在时刻 $n+1$ 由 $\frac{S_{n+1}}{S_n}$ 带来的随机性. 由自融资假定, 　此投资组合在时刻 $n+1$ 的价格应为

$$X_{n+1}=\Delta_nS_{n+1}-V_{n+1}=\Delta_nS_n\frac{S_{n+1}}{S_n}-V(n+1,S_n\frac{S_{n+1}}{S_n}). \tag{13. 40}$$

为了使后者不再随机，就是要使它的取值不再依赖于 a,b， 　也就应该有

$$\Delta_nS_na-V(n+1,aS_n)=\Delta_nS_nb-V(n+1,bS_n).$$

为了保证这个等式成立， 　我们只需把待定的 Δ_n 取为

$$\Delta_n=\frac{V(n+1,bS_n)-V(n+1,aS_n)}{(b-a)S_n}. \tag{13. 41}$$

利用无套利假定，这个消失了随机性的投资组合的时间发展应该按银行利率进行：

$$X_{n+1}=(1+R)X_n \tag{13. 42}$$

(如果取不等号就会如 Black-Scholes 模型时的推导那样出现套利). 于是我们有

$$X_n=\frac{1}{1+R}X_{n+1}=\frac{1}{1+R}(\Delta_nS_na-V(n+1,aS_n))$$

$$= \frac{1}{1+R}(\frac{a}{b-a}[V(n+1,bS_n)-V(n+1,aS_n)]-V(n+1,aS_n))$$
$$= \frac{1}{1+R}\frac{1}{b-a}[aV(n+1,bS_n)-bV(n+1,aS_n)]\text{．}$$

利用（１３．４１),（１３．４２）, 便得到

$$V(n,S_n)=\Delta_n S_n - X_n$$

$$= \frac{V(n+1,bS_n)-V(n+1,aS_n)}{b-a}-\frac{1}{1+R}\frac{1}{b-a}[aV(n+1,bS_n)-bV(n+1,aS_n)]$$
$$= \frac{1}{1+R}\frac{1}{b-a}[(1+R-a)V(n+1,bS_n)+(b-(1+R))V(n+1,aS_n)$$
$$= \frac{1}{1+R}[(1-p^*)V(n+1,bS_{n+1})+p^*V(n+1,aS_{n+1})]\text{．}$$

即函数$V(n,x)$**满足以下的后向差分方程：**

$$V(n,x)==\frac{1}{1+R}[(1-p^*)V(n+1,bx)+p^*V(n+1,ax)]. \qquad (13.\ 43)$$

并且由$V(N,S_N)=f(S_N)$**得到终端条件**

$$V(N,x)=f(x). \qquad (13.\ 44)$$

用数学归纳法便可得到（１３．４３）满足（１３．４４）的解为

$$V(n,x)=\frac{1}{(1+R)^{N-n}}\sum_{k=0}^{N-n} C_{N-n}^k p^{*k}(1-p^*)^{(N-n)-k} f(a^k b^{(N-n)-k}x)\ (n<N)\text{，}(13.45)$$

于是未定权益的贴水为

$$V(0,S_0)=\frac{1}{(1+R)^N}\sum_{k=0}^{N} C_N^k p^{*k}(1-p^*)^{N)-k} f(a^k b^{N-k}S_0).$$

在实际计算中， 可以用二叉树的后向传递法使用上述公式，并通过树上操作反向地逐步求出欧式未定权益在各个时刻的价格.

２．２ Black-Scholes **模型的二叉近似**

设风险证券的价格$S_t,(t\le T)$**由** Black-Scholes **模型描述．对于**$[0,T]$**作**m**等分．令**

$$S_n^{(m)} \overset{\Delta}{=} S_{\frac{nT}{m}}, \tilde{S}_n^{(m)} \overset{\Delta}{=} \frac{1}{(1+R)^n}S_n^{(m)}. \qquad (13.\ 46)$$

(**这时应该要求** $\tilde{S}_T=\tilde{S}_m^{(m)}$ **，即要求**$e^{-rT}S_T=\frac{1}{(1+R)^m}S_{Tt}$**，从而要求** $e^{-rT}=\frac{1}{(1+R)^m}$,

由此可由r,m**求出对应的**R**）．记**

$$\tilde{T}_n \overset{\Delta}{=} \frac{\tilde{S}_n^{(m)}}{\tilde{S}_{n-1}^{(m)}} = \frac{1}{1+R}\frac{S_n^{(m)}}{S_{n-1}^{(m)}}. \quad (13.47)$$

再假定 $\frac{S_n^{(m)}}{S_{n-1}^{(m)}}$ **服从风险中性的概率分布** $\begin{pmatrix} a & b \\ p^* & 1-p^* \end{pmatrix}$ （**其中** $p^* = \frac{b-(1+R)}{b-a}$ ）．**于是**

$\{\log \tilde{T}_n)$ **独立同分布，且**

$$\log \tilde{T}_n \sim \begin{pmatrix} \log\frac{a}{1+R} & \log\frac{b}{1+R} \\ p^* & 1-p^* \end{pmatrix}.$$

我们取

$$\log\frac{a}{1+R} = -s\sqrt{\frac{T}{m}}, \log\frac{b}{1+R} = s\sqrt{\frac{T}{m}}. \quad (13.48)$$

当 m **很大时有**

$$p^* = \frac{b-(1+R)}{b-a} = \frac{\frac{b}{1+R}-1}{\frac{b}{1+R}-\frac{a}{1+R}} = \frac{e^{s\sqrt{\frac{T}{m}}}-1}{e^{s\sqrt{\frac{T}{m}}}-e^{-s\sqrt{\frac{T}{m}}}}$$

$$\approx \frac{s\sqrt{\frac{T}{m}}+\frac{1}{2}s^2\frac{T}{m}}{2s\sqrt{\frac{T}{m}}} == (\frac{1}{2}+\frac{1}{4}s\sqrt{\frac{T}{m}}) \quad (13.49)$$

于是`

$$E(\log\tilde{T}_n) = p^*\log\frac{a}{1+R} + (1-p^*)\log\frac{b}{1+R}$$

$$\approx (\frac{1}{2}+\frac{1}{4}s\sqrt{\frac{T}{m}})(-s\sqrt{\frac{T}{m}}) + (\frac{1}{2}-\frac{1}{4}s\sqrt{\frac{T}{m}})(s\sqrt{\frac{T}{m}}) = -\frac{s^2}{2}\frac{T}{m} = o(1),(m\to\infty).$$

从而有

$$Var(\log\tilde{T}_n) \approx E(\log\tilde{T}_n)^2 = p^*(\log\frac{a}{1+R})^2 + (1-p^*)(\log\frac{b}{1+R})^2$$

$$\approx (\frac{1}{2}+\frac{1}{4}s\sqrt{\frac{T}{m}})s^2\frac{T}{m} + (\frac{1}{2}-\frac{1}{4}s\sqrt{\frac{T}{m}})s^2\frac{T}{m} = s^2\frac{T}{m}.$$

再则， 对于 $t \le T$ ，**由中心极限定理，当** m **大时得到** $\log\tilde{S}^{(m)}_{[\frac{t}{T/m}]} = \sum_{n=0}^{[\frac{t}{T/m}]}\log\tilde{T}_n$ **的分布与正态分布** $N(-\frac{s^2}{2}\frac{T}{m}[\frac{t}{T/m}], s^2\frac{T}{m}[\frac{t}{T/m}])$ **相近似，即与正态分布** $N(-\frac{s^2}{2}t, s^2t)$ **相近似．**

用更精细的讨论(用多维的中心极限定理)可以得到：随机过程 $\log \tilde{S}^{(m)}_{[\frac{t}{T/m}]}, (t \le T)$ 的有限维分布与漂移 Brown 运动 $sB_t - \frac{s^2}{2}t$ 有相同的有限维分布．也就是说，在 p^* 所对应的概率 $P^*(\cdot)$ 下，证券的折现价格 $\tilde{S}^{(m)}_{[\frac{t}{T/m}]}, (t \le T)$ 作为随机过程的有限维分布与几何 Brown 运动 $e^{sB_t - \frac{s^2}{2}t}$ 的有限维分布相同．而后者正是风险中性的 Black-Scholes 模型．这说明了离散二叉模型正是 Black-Scholes 模型的近似.

总之，对于给定的波动率 s 和利率 r 及充分大的 m，求出对应的离散利率 R，及二叉模型的 a, b 和公平概率 p^*，就可以用二叉模型 $\tilde{S}^{(m)}_{[\frac{t}{T/m}]}$ 来近似 Black-Scholes 模型．而前者的计算是很简单，可以用二叉树的后向传递法，通过树上操作求出未定权益在各个时刻的近似价格．

3．二叉模型的美式未定权益简述

３．１ 美式未定权益

我们不妨假定，标的证券在开始时刻的价格 S_0 是常数．

定义１３.１４ 如果容许证券的未定权益的持有人在 $[1, N]$ 中的任意时刻 n 都可以执行(但不在时刻 0 执行)，将持有人在时刻 n 执行此权益的所得记为 X_n，它是一个随机变量，一般是 $S_n, \cdots, S_1$ 的一个函数（即 X_n 是 (S_n) 可知的)． 与这种金融产品相系的权益序列 $\{X_n : n \le N\}$ 称为美式未定权益．

注意欧式未定权益就是 X_N，只是一个随机变量，而美式未定权益是一个随机变量序列．又因为美式未定权益可以在 $[1, N]$ 中的任意时刻执行，所以它的定价显然应该比欧式未定权益 X_N 的定价要高．在本节中，我们将对于二叉模型，给出美式未定权益的合理定价．而对于 Black-Scholes 模型，则可以与欧式未定权益一样，用二叉模型得到近似．

如果 $X_n = (S_n - K)^+$，则此未定权益称为美式看涨期权，同样地，如果 $X_n = (K - S_n)^+$，则此未定权益称为美式看跌期权．

对于二叉模型，$\frac{S_{n+1}}{S_n}$ 在风险中性的概率 P^* 下的分布为

$$\begin{pmatrix} a & b \\ p^* & q^* \end{pmatrix}, \quad p^* = \frac{b-(1+R)}{b-a}, \quad (q^* = 1 - p^*). \qquad (13.50)$$

记 S_n 的折现为

$$\tilde{S}_n = \frac{S_n}{(1+R)^n},$$

其中 R 为单位时间的银行利率．我们将美式未定权益 $\{X_n, n \le N\}$ 在时刻 n 的定价记为 U_n（它也是随机变量）．显见，在时刻 N 应有

$$U_N = X_N. \qquad (13．51)$$

而 $\{X_n, (n \le N)\}$ 在时刻 $N-1$ 的定价 U_{N-1}，则决定于当时执行的收益与在下一个时刻执行的预期收益中的较大者．由于它在下一时刻执行的预期收益应该是在时刻 N 到期的欧式未定权益在时刻 $N-1$ 的定价的折现为 $E^*(\frac{U_N}{(1+R)^N} | S_{N-1}, \cdots, S_1)$．于是

$$U_{N-1} = \max(X_{N-1}, (1+R)^{N-1} E^*(\frac{U_N}{(1+R)^N} | S_{N-1}, \cdots, S_1)). \qquad (13．52)$$

又因为 $U_N = X_N$，上式即为

$$U_{N-1} = \max(X_{N-1}, (1+R)^{N-1} E^*(\frac{X_N}{(1+R)^N} | S_{N-1}, \cdots, S_1)). \qquad (13．52)'$$

它可以改写为

$$U_{N-1} = (1+R)^{N-1} \max_{m \ge N-1} E^*(\frac{X_m}{(1+R)^m} | S_{m-1}, \cdots, S_1). \qquad (13．52)''$$

一般地，美式未定权益在时刻 n 的定价，应该决定于当时执行的收益与把下一个时刻定价 U_{n+1} 作为时刻 $n+1$ 到期的欧式未定权益的定价中的较大者．由于在下一时刻到期的欧式未定权益 U_{n+1} 的定价的折现应为 $E^*(\frac{U_{n+1}}{(1+R)^{n+1}} | S_n, \cdots, S_1)$．于是我们有

$$U_n = \max(X_n, (1+R)^n E^*(\frac{U_{n+1}}{(1+R)^{n+1}} | S_n, \cdots, S_1)). \qquad (13．53)$$

利用条件期望的性质及 $U_n \ge X_n$，用归纳法不难证明

$$U_n = (1+R)^n \max_{m \ge n}(E^*(\frac{X_m}{(1+R)^m} | S_n, \cdots, S_1)), \qquad (13．54)$$

$$U_0 = \max_{m \geq 0} E^*(\frac{X_m}{(1+R)^m}) \qquad (13.54)'$$

命题１３．１４ 美式未定权益的折现定价$\tilde{U}_n$ ($\tilde{U}_n = \frac{U_n}{(1+R)^n}, (n \leq N)$)是$(S_n)$上鞅列，而且它是满足$U_n \geq X_n$条件的上鞅列中的最小的一个，所以$\tilde{U}_n$称为控制$\tilde{X}_n$的最小上鞅列，也称为$\tilde{X}_n$的 Snell 包络，其中$\tilde{X}_n = \frac{X_n}{(1+R)^n}$为未定权益的折现.

证明 前一个结论直接得自(13．53)．又若Y_n是另一个控制$\tilde{X}_n$的上鞅列，我们用后向归纳法来证明$Y_n \geq \tilde{U}_n$．对于$n = N$，我们有$Y_N \geq \tilde{X}_N = \tilde{U}_N$．假定$n = m$时有$Y_m \geq \tilde{U}_m$，那么

$$Y_{m-1} \geq E^*(Y_m \mid S_{m-1}, \cdots, S_1) \geq E^*(\tilde{U}_m \mid S_{m-1}, \cdots, S_1).$$

再利用(13．53)便得$Y_{m-1} \geq \tilde{U}_{m-1}$．由归纳法推出结论成立．】

更一般地，对于任意一个取值于$[n, N]$的(S_k)停时t，对它的不同可能取值用全期望公式还可以得到

$$U_n = (1+R)^n \max_{\text{一切停时 } t \in [n,M]} E^*[\frac{X_t}{(1+R)^t} \mid S_n, \cdots, S_1]. \qquad (13.55)$$

所以，美式未定权益在初始时刻的贴水(保证金)又可以表示为

$$U_0 = \max_{\text{一切停时 } t \in [0,M]} E^*[\frac{X_t}{(1+R)^t}]. \qquad (13.56)$$

等式(13．55)与(13．56)并不能用于实算．但是，我们可以用它来讨论未定权益的持有人在理论上于什么时刻执行最有利，这种有利时刻称为最佳执行时刻．

定义１３.１６ （未定权益的最佳执行时刻） 任意一个(S_k)停时t^*，只要满足

$$U_0 = E^*[\frac{X_{t^*}}{(1+R)^{t^*}}] \ (= \max_{\text{一切停时 } t \in [0,M]} E^*[\frac{X_t}{(1+R)^t}]),$$

就称t^*为美式未定权益的一个最佳执行时刻，其中最后一个等号正说明了它的最佳性质．由此定义可以看到，最佳执行时刻并不一定唯一．

在一般时刻n，(13．53)指出了定价$U_n \geq$(该时刻执行的所得权益)X_n．直观地，再对未定权益的持有人有利的时刻执行所得也不应超过定价．但是在定价随机变量与执行所

得的权益随机变量相等的那些时刻(一般都是随机时刻)上，执行所得收益就达到了它的上界．因此，这些随机时刻就应是最佳执行时刻．于是我们有下面的定理．

定理１３．１７ 把定价与未定权益首次相等的时刻(是一个停时)记为t_0，即

$$t_0 = \min\{ n > 0 : U_n = X_n \}．$$

那么，它也是美式未定权益的一个最佳执行时刻．

(此定理在直观上是显然的，但是其证明则需要用鞅列在停时上的停止定理，本书从略)．

［注］ 定价的折现$\tilde{U}_n$是(S_n)上鞅列，这就是说，由预期美式未定权益的收益而作出的定价平均地是下降的．它的直观意义很清楚，时间越早对持有人有更大的选择余地，定价当然就应该高一些．所以，美式未定权益的定价作为随机变量(按上鞅的规律发展)，其平均值是时间的下降序列．这一点正体现了定价在平均的意义上，对于买方与卖方的公平性．

再则，即使持有人在最佳执行时刻执行，所得到收益的折现也不过就是当时的定价．那么，购买这种未定权益能有那么好处呢 ？事实上，虽然定价在平均意义上不过是一种公平的博采，但是正如许多公平的博采一样，人们还是抱着试试运气的心理参加投资，希望在个别的情形中，得到的报酬会远高于平均所得．显然，在别的一些情形，投资人也可能得到远低于平均的报酬．那么，为什么这种金融工具会广泛流行呢 ？ 事实上，对于全社会而言，这样的金融工具是一种融资的手段，能分担社会的金融风险，促进资本有效流动．当然，这并不包括黑户操纵与违规操作．而对投资人来说，在健全的金融体制下，虽然有人会得利，也有人会受到损失，只要局限在经济力量容许的范围内，利用闲散资金投资，也是对金融的某种参与，况且在正常的情形,标的证券还可以得到分红．当然，任何金融创新都是一把双刃剑，在起正面作用的时候，又常常被利用作为投机手段，这可能带来另一种新的风险．这是人类社会的普遍的客观规律．

需要特别注意的是，实际市场的炒作与风险，远比数学模型复杂．数学模型只是平均行为的一个非常粗放的参考，哪种模型近似得较为合理一些，应该以实证数据为根据．

3．2 二叉模型美式未定权益$\{f(S_n), n \le N\}$的定价与定价函数组

本段我们讨论二叉模型美式未定权益的一种常见的特殊情形，即存在$f(x)$使对于任意$n \le N$，都有

$$X_n = f(S_n)$$

的情形(典型情形出现于美式看涨期权和美式看跌期权，它们分别对应于$f(x) = (x-K)^+$和$f(x) = (K-x)^+$)，此时，由(13．51)，(13．52)和(13．53)，可以直接推出

$$U_N = f(S_N) \overset{\Delta}{=} U_N(S_N)，(其中\ U_N(x) = f(x))， \qquad (13．57)$$

$$U_{N-1} = \max\left(f(S_{N-1}), \frac{1}{1+R} E^*(f(S_N) \mid S_{N-1}, \cdots, S_1)\right)$$

$$= \max\left(f(S_{N-1}), \frac{1}{1+R}[p^* f(aS_{N-1}) + q^* f(bS_{N-1})]\right) \overset{记为}{=} U_{N-1}(S_{N-1})，$$

其中

$$U_{N-1}(x) = \max(f(x), \frac{1}{1+R}[p^* f(ax)+q^* f(bx)]) .$$

对于一般$n<N$，用后向归纳法可得

$$U_n = \max(f(S_n), \frac{1}{1+R}[p^*U_{n+1}(aS_n)+q^*U_{n+1}(bS_n)]) \overset{\text{记为}}{=} U_n(S_n) ,$$

其中

$$U_n(x) = \max(f(x), \frac{1}{1+R}[p^*U_{n+1}(ax)+q^*U_{n+1}(bx)]) . \qquad (13．58)$$

由此我们归纳出下面的定理

定理１３．１８ 只要后向地逐个从以下的终端已知的极值函数方程组：

$$U_N(x) = f(x) ,$$

$$U_n(x) = \max(f(x), \frac{1}{1+R}[p^*U_{n+1}(ax)+q^*U_{n+1}(bx)]) , (n<N) \qquad (13.59)$$

解出$\{U_n(x): n\le N\}$，便得到美式未定权益的定价 $U_n = U_n(S_n)$.

定义１３．１９ （美式未定权益的定价函数组） 由定理１３．１８所确定的$\{U_n(x): n\le N\}$，称为美式未定权益$\{f(S_n)\}$ $(n\le N)$的定价函数组.

于是对于二叉模型，美式未定权益的定价可以由极值函数方程组的数值求解得到． 即为了达到最佳执行，只须用下述算法：

(1) 对于$f(x)$，计算出定价函数组 $\{U_n(x): n\le N\}$；

(2) 对于证券的实时市场价 $S_0, S_1,\cdots$ ，实时地计算 $f(S_1), f(S_2),\cdots$ 与 $U_1(S_1), U_2(S_2),\cdots$，并实时地依次比较对应的值；

(3) 在计算(2)的过程中，一旦发现它们在某个时刻相等，则此时刻就是最佳执行时刻．即：若样本列 $f(S_1), f(S_2),\cdots$ 与定价列$U_1(S_1), U_2(S_2),\cdots$第一次相等出现于$m$，则未定权益的持有人就在此时刻$m$执行.

美式未定权益的套期与消费过程

由于折现价格序列$\tilde{U}_n$是上鞅列，于是按如下定义的随机变量列$\tilde{C}_n$为非负的：

$$\tilde{C}_n = \tilde{U}_n - E^*(\tilde{U}_{n+1} \mid S_n,\cdots,S_1) \ge 0 . \qquad (13．60)$$

又因为美式未定权益在任意时刻都可以执行，我们只能仿照欧式未定权益的套期公式来套期．也就是说，在时刻n的套期标的证券的数目取为

$$\Delta_n = \frac{U_{n+1}(bS_n) - U_{n+1}(aS_n)}{(b-a)S_n} , \qquad (13．61)$$

此处与欧式未定权益的情形唯一的不同之处是，用了美式未定权益的定价$\{U_n\}$代替了欧式未定权益的定价$\{V_n\}$．令 $C_n=(1+R)^n\tilde{C}_n$，于是

$$C_n=U_n-\frac{1}{1+R}E^*(U_{n+1}\mid S_n,\cdots,S_1)\,. \tag{13．62}$$

由于

$$E^*(U_{n+1}\mid S_n,\cdots,S_1)=p^*U_{n+1}(aS_n)+q^*U_{n+1}(bS_n)\,,$$

我们有

$$U_{n+1}(aS_n)=q^*[U_{n+1}(aS_n)-U_{n+1}(bS_n)]+p^*U_{n+1}(aS_n)+q^*U_{n+1}(bS_n)$$

$$=-[(1+R)-a]\Delta_nS_n+E^*(U_{n+1}\mid S_n,\cdots,S_1)$$

$$=-(1+R)\Delta_nS_n+\Delta_n\cdot aS_n-(1+R)(C_n-U_n)\,.$$

完全类似地还可以证明

$$U_{n+1}(bS_n)=-(1+R)\Delta_nS_n+\Delta_n\cdot bS_n-(1+R)(C_n-U_n)\,.$$

把$S_{n+1}=aS_n$与$S_{n+1}=bS_n$两种可能的情形合起来，就得到

$$U_{n+1}=\Delta_nS_{n+1}+(1+R)(U_n-C_n-\Delta_nS_n)\,.$$

把它改写为更有金融含义的如下等式

$$(U_{n+1}-\Delta_nS_{n+1})=(1+R)(U_n-\Delta_nS_n)-(1+R)C_n\,. \tag{13．63}$$

这公式具有比较清楚的金融含义，它说明在美式未定权益情形，即使用合适的套期Δ_n，也不能使余下的资产$U_n-\Delta_nS_n$为无风险资产．只有在时刻n的套期余额中消耗了C_n以后的资产，到下一个时刻$n+1$的价值才是时刻$n+1$的套期后的余额．这个随机序列$\{C_n\}$称为美式未定权益的消费过程． 一般地，不管是什么模型，美式未定权益的套期总需要有一个消费过程去补充．在理论上知道美式未定权益的定价需要补充一个消费过程，是对美式未定权益定价模型认识的一个实质性的进展．

4．随机利率与债券利率的期限结构

4．1　$s-$零息债券

定义１３．２０　不付红利且在时刻s到期的债券称为$s-$零息(Zero-Coupon)债券．

$s-$零息债券进入市场后，在市场操作下，相当于按某个随机利率增值．记时刻$t(<s)$的 1 元钱按市场发展在债券到期时刻s的价值(“本利和”)为$R(t,s)$．把时刻s的 1 元钱

在时刻$t(<s)$的定价记为$P(t,s)$．显见

$$P(t,s)=\frac{1}{R(t,s)},P(s,s)=1. \qquad (13．64)$$

一般地，$P(t,s)$是随机过程．注意，由于$s-$零息债券与$t-$零息债券在市场的表现是不一样的，所以即使在无套利的假定下，一般地

$$P(0,s)\neq P(0,t)P(t,s),(t<s). \qquad (13．65)$$

事实上，这是因为(13．65)的左边是由初始时刻的市场信息所确定，而右边是由直至时刻t为止的信息所确定的，因此一般不会相等．

但是如果定价$\{P(t,s),0\le t\le s)\}$不是随机过程而是确定性函数时，对于债券而言，在时刻t与在初始时刻有相同的信息．此时在无套利假定下就有

$$P(0,s)=P(0,t)P(t,s),(t<s).$$

4．2　零息债券导出的各种的随机利率概念

定义１３．２１（平均远期利率（到期收益率））　记

$$\overline{f}(t,s)=-\frac{\log P(t,s)}{s-t},(t<s), \qquad (13．66)$$

即

$$P(t,s)=e^{-\overline{f}(t,s)(s-t)},(t<s). \qquad (13．66)'$$

$\overline{f}(t,s)$称为$s-$零息债券的平均远期利率．此处的 ”远”的含义是指参考时间$t>0$，即不是初始时刻的利率，它是在将来的时刻$t(<s)$，展望$s-$债券的平均利率．　由$1=P(t,s)e^{\overline{f}(t,s)(s-t)},(t<s)$说明平均远期利率就是在时刻$t$把$s-$债券换为比它早到期的$t-$债券所需付的利率．

在时刻$t(<s)$固定时，平均远期利率作为$s-t$的函数的图形变化，称为利率的期限结构．

更一般地，有下面的定义．

定义１３．２２　对于涉及不同的到期时刻的两种零息债券，定义

$$\overline{f}(t,s_1,s_2)=-\frac{\log P(t,s_2)-\log P(t,s_1)}{s_2-s_1}，(0<t<s_1<s_2)\ ， \qquad (13．67)$$

即

$$P(t,s_1)=P(t,s_2)e^{\overline{f}(t,s_1,s_2)(s_2-s_1)}.$$

这说明$\overline{f}(t,s_1,s_2)$就是在时刻t把s_2-债券换为比它早到期的s_1-债券所需付的利率．

按定义有

$$\overline{f}(t,s)=\overline{f}(t,t,s)\,. \qquad (13.68)$$

定义 13.23 （平均即期利率，Spot Rate） $\overline{f}(0,s)$ **称为平均即期利率．**

（于是我们有 $P(0,s)=P(0,0)e^{-s\overline{f}(0,s)}$**，此即(13.66)’的特殊情形）．**

定义 13.24 （远期利率，Instantaneous Forward Rate） **若** $P(t,s)$ **对** s **可微分，则**

$$f(t,s)\overset{\Delta}{=}-\frac{\partial \log P(t,s)}{\partial s}(=\lim_{\Delta s\to 0}\overline{f}(t,s,s+\Delta s)) \qquad (13.69)$$

称为远期利率．于是有

$$P(t,s)=e^{-\int_t^s f(t,u)du}, \qquad (13.70)$$

$$\overline{f}(t,s)=\frac{1}{s-t}\int_t^s f(t,u)du, \qquad (13.71)$$

((13．71)式也可证明如下 $\frac{d}{ds}[(s-t)\overline{f}(t,s)]=-\frac{d\log P(t,s)}{ds}=f(t,s)$**）．由(13.70)式与(13.71)式可以看出，给出定价随机过程** $\{P(t,s),0\le t\le s)\}$ **与给出远期利率随机过程** $\{f(t,s),0\le t\le s)\}$ **是等效的．**

定义 13.25 （短期利率 Short Rate ） $r(t)\overset{\Delta}{=}f(t,t)$ **称为短期利率．**
它就是普通习惯上所称的利率，不过它是随机的．我们有

$$r(t)=f(t,t)=\overline{f}(t,t)\,. \qquad (13.72)$$

（证明 $f(t,t)$ **与** $\overline{f}(t,t)$ **按定义都是** $\lim_{s\downarrow t}-\frac{\log P(t,t+(s-t))}{s-t}$**）．**

例 １３．２６（非随机情形） **此时** $\{P(t,s),0\le t\le s)\}$ **是确定性函数，因而所涉及的各种利率也都是确定性的函数．在不十分剧烈炒作的场合下，可以用此看法对实际市场作近似．此时各种利率可以有如下的一些换算公式：对于** $0\le t\le s$ **有**

(1) $P(0,s)=P(0,t)P(t,s),(t<s)$**．**

(2) $f(0,s)=f(t,s)$**，即** $f(t,s)$ **不依赖** t**．从而** $f(t,s)=f(s,s)=r(s)$**．**

(3) $\overline{f}(t,s)=\frac{1}{s-t}\int_t^s r(u)du$**，从而有** $P(t,s)=e^{-\int_t^s r(u)du}$ **．**

(4) $\overline{f}(t,s)=\dfrac{s\overline{f}(0,s)-t\overline{f}(0,t)}{s-t}$. (13.73)

一般从交易所的挂牌数据，可以得到的平均即期利率 $\overline{f}(0,t),\overline{f}(0,s)$，再用(13.73)式算出平均远期利率 $\overline{f}(t,s)$． 而在实际交易中，人们用的正是平均远期利率．由(13.73)式还可以得到

$$\overline{f}(0,s)\geq\overline{f}(0,t)\Rightarrow\overline{f}(t,s)\geq\overline{f}(0,s)$$

(5) $r(t)=\overline{f}(0,t)+t\dfrac{d}{dt}\overline{f}(0,t)$.

(证明　将(4)改写为 $\overline{f}(t,s)=\overline{f}(0,s)+(\overline{f}(0,s)-\overline{f}(0,t))\dfrac{t}{s-t}$. 令 $s\downarrow t$，便得(5))．

4．3　资产定价基本定理与利率衍生证券

假定时刻 t 的短期利率 $r(t)$ 可由市场在 t 以前的信息，例如 $(B_t^*:u\leq t)$，所完全确定，其中 B_t^* 是 Brown 运动．此时我们有下面的定理．

定理１３．２７　（资产定价基本定理 ）

市场无套利性等价于具有风险中心概率 P^*，使得在此概率下，折现值 $\tilde{P}_t\stackrel{\Delta}{=}\dfrac{P(t,s)}{e^{\int_0^t r(u)du}}$ 为 (B_t^*) 鞅，其中 $P(t,s)$ 是 1 元的 $s-$ 零息债券在时刻 $t(<s)$ 的价格．于是

$$P(t,s)=e^{\int_0^t r(u)du}E^*(\tilde{P}_s\mid B_u^*:u\leq t).$$

即 $s-$ 零息债券的无套利定价为

$$P(t,s)=E^*(e^{-\int_t^s r(u)du}\mid B_u^*:u\leq t). \qquad (13.\ 74)$$

反之，(13.７４)也保证了 $\tilde{P}_t\stackrel{\Delta}{=}\dfrac{P(t,s)}{e^{\int_0^t r(u)du}}$ 是 (B_t^*) 鞅.

由此看出，只要给定了短期利率 $r(t)$ 的风险中性模型，就可以得到 $s-$ 零息债券的无套利定价．

定义１３．２８　一个金融合约的权益，如果依赖于未来的利率，或债券的价格，则此合约就称为利率未定权益，或利率衍生证券．

随机利率模型的研究，是要给出 $s-$ 债券的定价，远期利率及利率未定权益的定价．

4．4．利率的风险中性模型

一般地，只要给定$r(t)$，由(13．74)定义的$P(t,s)$就保证了$\tilde{P}_t \overset{\Delta}{=} \frac{P(t,s)}{e^{\int_0^t r(u)du}}$为$(B_t^*)$鞅，从而$P(t,s)$就是无套利定价，此时用以确定$r(t)$的随机微分方程的模型，就是风险中性模型．所以短期利率的风险中性模型可以任意给定．不带什么约束条件（事实上，从(13．74)求出的远期利率为$f(t,s) = E^*(r(s)e^{-\int_t^s r(u)du} \mid B_u^* : u \le t)$，得到的$f(t,t)$正好等于$r(t)$）．与此相反的是，如果直接给出价格$P(t,s)$满足随机微分方程，或者等价地，直接给出远期利率$f(t,s)$满足的随机微分方程，由此确定的$r(t)$，就未必满足(13．74)．这时，(13．74)就变为 约束条件．只有满足了这个约束条件的$P(t,s)$才是无套利定价，所对应的$P(t,s)$或$f(t,s)$的随机微分方程模型才是风险中性模型．

远期利率的 Heath-Jarrow-Morton 模型及其风险中性情形

由资产定价基本定理可以知道，为了使模型成为风险中性，充要条件是$\tilde{P}_t \overset{\Delta}{=} \frac{P(t,s)}{e^{\int_0^t r(u)du}}$是鞅．由此利用(13．70)及 Girsanov 定理，就能推出以下的定理．

定理１２．２９　（Heath-Jarrow-Morton 定理）　对于固定的s，如果远期利率$f(t,s)$是以下的 Ito 过程

$$df(t,s) = b(t,s)dt + s(t,s)dB_t\,, \qquad (13．75)$$

其中$b(t,s), s(t,s)$在s固定下是(B_t)可知的随机过程．那么，此模型是无套利(即风险中性的)的充要条件为：存在一个满足条件

$$Ee^{-\int_0^s q(t)dB_t - \frac{1}{2}\int_0^s q(t)^2 dt} = 1$$

的(B_t)可知的随机过程$q(t)$，使

$$b(t,s) = s(t,s)[\int_t^s s(t,u)du + q(t)]\,.$$

在条件成立时，由 Girsanov 定理可知，$B_t^* \overset{\Delta}{=} B_t + \int_0^t q(u)du$也是 Brown 运动，且

$$df(t,s) = s(t,s)[\int_t^s s(t,u)du]dt + s(t,s)dB_t^*\,. \qquad (13．76)$$

再对 $P(t,s)=e^{-\int_t^s f(t,u)du}$ 用 Ito 公式可得

$$dP(t,s)=P(t,s)[r(t)dt-(\int_t^s s\,(t,u)du)dB_t^*]\,. \qquad (13.77)$$

可见在此情形下，$s-$零息债券的价格 $P(t,s)$ 的收益率为 $r(t)$，波动率为 $\int_t^s s\,(t,u)du$ (而

$s\,(t,s)/f\,(t,s)$ 为远期利率的波动率).

直接给出短期利率风险中性模型的有

(1) Vasicek 模型

假定短期利率 $r(t)$ 满足如下的 Ornstein-Uhlenbeck 随机微分方程:

$$dr(t)=a(b-r(t))dt+s\cdot dB_t^*\,. \qquad (13.78)$$

这个随机微分方程具有显式解:

$$r(t)=r(0)e^{-at}+b(1-e^{-at})+s\cdot e^{-at}\int_0^t e^{as}dB_s^*\,. \qquad (13.79)$$

容易证明它是 Gauss 过程. 而作为 Gauss 过程的积分的 $\int_0^t r(u)du$ 也是 Gauss 过程. 于是

可以仿照风险中性概率推导 Black-Scholes 公式的计算过程算出(13.74)中的条件期望. 经过仔细而并不复杂的推导计算，便可得到

$$P(t,s)=e^{-A(t,s)-C(t,s)r(t)}, \qquad (13.80)$$

其中

$$A(t,s)=\frac{1}{a^2}(ts-C(t,s))(a^2b-\frac{s^2}{2})+\frac{s^2C(t,s)^2}{4a}, \quad C(t,s)=\frac{1-e^{-a(s-t)}}{a}\,.$$

于是远期利率为

$$f\,(t,s)=\frac{C(t,s)r(t)+A(t,s)}{s-t},$$

它是短期利率的线性函数.

(2) Hull-White 模型

Hull-White 模型是 Vasicek 模型在时变系数情形的推广，即模型参数 a,b,s 分别被推

广为确定性的函数 $a(t),b(t),s\,(t)$

$$dr(t)=a(t)(b(t)-r(t))dt+s\cdot(t)dB_t^*\,. \qquad (13.78)'$$

与常数情形完全类似地可以算出定价过程仍是 $r(t)$ 的如下函数

$$P(t,s)=e^{-A(t,s)-C(t,s)r(t)}，\qquad (13.80)'$$

只需要对 $A(t,s),C(t,s)$ 作一些相应的变更.

(13.80)'说明，此时 $s-$ 零息债券的价格 $P(t,s),(t\le s)$ 是 $r(t)$ 的函数．由于 $r(t)$ 是 Ito 过程,．用 Ito 公式就得到

$$dP(t,s)=P(t,s)[r(t)dt-s\ (t)C(t,s)dB_t^*]\ .\qquad (13.81)$$

这就是说，$P(t,s)$ 的收益率是 $r(t)$，波动率是 $s\ (t)C(t,s)$．(13.80)'还可以用以从实证数据校正模型参数函数组 $a(t),b(t),s\ (t)$．Hull-White 模型的缺点是，短期利率服从正态分布，因而可以取负值，这是不合理的.

Hull-White 模型下的债券看涨期权

对于时刻 $T(<s)$ 到期的，以 $s-$ 零息债券为标的证券的，且执行价格为 K 的欧式看涨期权，其未定权益为 $[(P(T,s)-K]^+$．于是它在时刻 t 的价格应该为

$$V_t\overset{\Delta}{=}E^*(e^{-\int_t^s r(u)du}[P(T,s)-K]^+\mid B_u^*:u\le t)\ .\qquad (13.81)$$

利用 $(r(t),\int_0^t r(u)du)$ 是 Gauss 过程，就可以计算出这种期权的贴水 V_0 的明显公式.

(3) CIR (Cox-Ingersoll-Ross) 模型

Cox-Ingersoll-Ross 假定短期利率 $r(t)$ 满足下述随机微分方程

$$dr(t)=a(b-r(t))dt+s\ \cdot\sqrt{r(t)}dB_t^*.\qquad (13.82)$$

(如果有 $d(\ge 2)$ 个彼此独立的 Brown 运动 $B_t^{(1)},\cdots,B_t^{(d)}$，又对于任意 $i\le d$ 有

$$dX_t^{(i)}=-\frac{a}{2}X_t^{(i)}dt+s\ \cdot dB_t^{(i)}\ .$$

那么对于 $r(t)\overset{\Delta}{=}\sqrt{(X_t^{(1)})^2+\cdots+(X_t^{(d)})^2}$，可以证明存在 Brown 运动 B_t^* 使

$$dr(t)=(\frac{s^{\ 2}d}{4}-a\cdot r(t))dt+s\ \cdot\sqrt{r(t)}dB_t^*\ .$$

这正是 CIR 模型)．这时 $r(t)$ 是扩散过程,求解不变密度的方程

$$\frac{d^2}{dr^2}(\frac{s^{\ 2}r}{2}j\ )-\frac{d}{dr}(a(b-r)j\ )=0$$

就得到不变密度

$$j(r) = Cr^{\frac{2ab-s^2}{s^2}} e^{-\frac{2a}{s^2}r}, (r>0)\text{，} \qquad (13.83)$$

其中

$$C = \left(\frac{2a}{s^2}\right)^{\frac{2ab}{s^2}} \frac{1}{\Gamma\left(\frac{2ab}{s^2}\right)}\text{.}$$

CIR 模型的债券的定价

此时$\{B_u^* : u \le t\}$与$\{r(u) : u \le t\}$具有相同的信息，所以

$$P(t,s) = E^*(e^{-\int_t^s r(u)du} \mid B_u^* : u \le t) = E^*(e^{-\int_t^s r(u)du} \mid r(u) : u \le t)\text{.}$$

由$r(t)$的 Markov 性可知上式右边的量只依赖于$r(t)$，即存在一个数值函数$F(t,r,s)$使

$$P(t,s) = F(t, r(t), s)\text{.} \qquad (13.84)$$

假定$F(t,r,s)$对r二阶连续可微，对t连续可微， 那么$e^{\int_0^t r(u)du} F(t,r(t),s)$是一个 Ito 过程．利用推广的 Ito 公式可算出

$$d(e^{\int_0^t r(u)du} F(t,r(t),s)) = [\frac{\partial F}{\partial t} + \frac{s^2 r(t)}{2}\frac{\partial^2 F}{\partial r^2} + a(b-r(t))\frac{\partial F}{\partial r} - r(t)F]dt + s \cdot \sqrt{r(t)}\frac{\partial F}{\partial r} dB_t^*$$

另一方面，由定义$e^{\int_0^t r(u)du} F(t,r(t),s) = E^*(e^{-\int_0^s r(u)du} \mid B_u^* : u \le t)$是一个$(B_t^*)$鞅，两者比较可知$F$必须满足偏微分方程：

$$\frac{\partial F}{\partial t} + \frac{s^2 r}{2}\frac{\partial^2 F}{\partial r^2} + a(b-r)\frac{\partial F}{\partial r} - rF = 0\text{.} \qquad (13.85)$$

及终端条件

$$F(s,r,s) = 1\text{.} \qquad (13.86)$$

我们通过待定$A(t,s), C(t,s)$，探求方程的以下形式的解的可能性

$$F(t,r,s) = e^{-A(t,s) - C(t,s)r}\text{.} \qquad (13.87)$$

将(13.87)代入（13.85)，就得到$A(t,s), C(t,s)$关于t的二阶常微分方程．求解得到

$$C(t,s) = \frac{sh(g(s-t))}{g \cdot ch(g(s-t)) + \frac{a}{2} sh(g(s-t))}\text{，}$$

$$A(t,s)=-\frac{2ab}{s^2}\log\frac{g\cdot e^{\frac{a}{2}(s-t)}}{g\cdot ch(g(s-t))+\frac{a}{2}sh(g(s-t))},$$

其中

$$g=\frac{1}{2}\sqrt{a^2+2s^2},$$

而 $ch(x), sh(x)$ 分别是双曲余弦与双曲正弦：

$$ch(x)=\frac{e^x+e^{-x}}{2},\quad sh(x)=\frac{e^x-e^{-x}}{2}.$$

CIR 模型的债券看涨期权的定价

对于时刻 $T(<s)$ 到期的，以 $s-$ 零息债券为标的证券的，且执行价格为 K 的欧式看涨期权，其未定权益为 $[(P(T,s)-K]^+$．于是它在时刻 t 的价格应该为

$$V_t\overset{\Delta}{=}E^*(e^{-\int_t^s r(u)du}[P(T,s)-K]^+ \mid B_u^*:u\le t)$$

$$=E^*(e^{-\int_t^s r(u)du}[F(T,r(T),s)-K]^+ \mid B_u^*:u\le t). \qquad (13.88)$$

与债券的定价类似地，可假定存在数值函数 $G(t,r,T,s)$ 使债券看涨期权的价格 $V_t=G(t,r(t),T,s),(t\le T<s)$．再利用

$e^{\int_0^t r(u)du}G(t,r(t),T,s)=E^*(e^{-\int_0^s r(u)du}(P(T,s)-K)^+ \mid B_u^*:u\le t)$ 是一个 (B_t^*) 鞅，就得到 $G(t,r,T,s)$ 满足的偏微分方程

$$\frac{\partial G}{\partial t}+\frac{s^2r}{2}\frac{\partial^2G}{\partial r^2}+a(b-r)\frac{\partial G}{\partial r}-rG=0,(t\le T), \qquad (13.89)$$

及终端条件

$$G(T,r,T,s)=(F(T,r,s)-K)^+. \qquad (13.90)$$

*** 5．基于证券的随机利率的债券为币值单位折现的证券的未定权益的定价**

设证券价格 S_t 与债券价格 b_t 分别满足随机微分方程

$$dS_t=m(t,S_t)dt+s(t,S_t)dB_t,$$

$$db_t = r(t, S_t) b_t dt \text{ ,}$$

其中 B_t 为 Brown 运动，而债券 b_t 的市价是以依赖于证券值 S_t 的短期的连续随机利率 $r = r(t, S_t)$ 而增值的．与 Black-Scholes 模型类似地，将 $v(t, S_t) = \dfrac{m(t, S_t) - r(t, S_t)}{s(t, S_t)}$ 称为证券 S_t 的风险的市场价格．与 Black-Scholes 模型不同处是，这里不是用按非随机的利率增值的银行存款作为证券币制单位，而是按用与证券密切相关的随机利率增值的债券 b_t 作为该证券的币值单位．

于是，与 Black-Scholes 模型类似地,如下确定的新的概率

$$P^*(A) = E(e^{-\int_0^T v(t, S_t) dB_s - \frac{1}{2}\int_0^T v)t, S_t)^2 ds} I_A)$$

称为风险中性概率．在此概率下证券按债券的折现价格 $\tilde{S}_t \overset{\Delta}{=} \dfrac{S_t}{b_t}$ 是 (S_t) 鞅．

对此模型的未定权益 $f(S_T)$ 在时刻 t 的定价为

$$F(t, S_t) = E^*[f(S_T)\frac{b_t}{b_T} | S_t, \cdots, S_1] \text{ .}$$

同样，与 Black-Scholes 模型类似地，价格函数 $F(T, x)$ 应满足如下的带终端条件的偏微分方程：

$$F_t + \frac{s^2(t, x)}{2} x^2 F_{xx} + r(t, x) x F_x - r(t, x) F = 0 \text{ ,}$$

$$F(T, x) = f(x) \text{ .}$$

在群体保险中的应用

设年龄 x 参加保险的一个群体共有 n 人，假定各人的余寿是独立同分布的随机变量 $T_i^x, (i \le n)$．又设个体在年龄 $x + t$ 时的死亡率为 m_{x+t} ，即

$$G(t) \overset{\Delta}{=} P(T_i^x \ge t) = e^{-\int_0^t m_{x+s} ds} \text{ .}$$

令 N_t 为此群体中在时刻 t 以前死亡的人数(死亡计数过程)，即

$$N_t = \sum_{i=1}^n I_{\{T_t^x \le t\}} \text{ .}$$

于是有

$$EN_t = n[1 - G(t)] \text{ ,}$$

$$E[(N_{t+\Delta t}-N_t)]=\sum_{i=1}^{n}[(1-G(t+\Delta t))-(1-G(t))]=n[G(t)-G(t+\Delta t)]$$

$$=-ne^{-\int_0^t m_{x+s}ds}(1-e^{-\int_t^{t+\Delta t} m_{x+s}ds})=-(n-EN_t)m_{x+t}(\Delta t+o(\Delta t)).$$

注意到各人的生死是彼此独立的，故而有

$$E[(N_{t+\Delta t}-N_t)\mid N_s:s\le t]=(n-N_t)m_{x+t}[\Delta t+o(\Delta t)]\overset{\Delta}{=}l_t[\Delta t+o(\Delta t)],$$

其中

$$l_t=(n-N_t)m_{x+t}=\lim_{\Delta t\to 0}\frac{1}{\Delta t}E[(N_{t+\Delta t}-N_t)\mid N_s:s\le t]$$

是一个随机过程，它代表此群体在时刻 t 的死亡强度，也称为在时刻 t 的集体死亡风险率．这时

$$\tilde{N}_t\overset{\Delta}{=}N_t-\int_0^t l_s ds$$

是 (N_t) 鞅．随机过程 $\int_0^t l_s ds$ 称为计数过程 N_t 的补偿过程．

这种模型的典型是养老保险合约 (Pure Endowment)． 按照这种合约，如果投保人在保险到期时刻 T 还未死亡，则可以从保险公司领到金额 $f(S_T)$．于是公司对此项保险的群体应付金额在起始时刻的折现值为

$$\tilde{H}=\sum_{i=1}^{n}I_{\{T_i>T\}}f(S_T)b_T^{-1}=(n-N_T)f(S_T)b_T^{-1}.$$

于是，这个折现了的未定权益在时刻 t 的折现定价为

$$\tilde{V}_t=E^*(\tilde{H}\mid S_u,N_u:u\le t),$$

其中 P^* 为风险中性概率．要计算这个条件期望，或者要得到它所满足的决定性的方程，需要用更多的随机分析知识，包括一般的所谓半鞅的积分及更为一般的 Ito 公式．

习题 13

1． 证明直接求解随机微分方程 $d\tilde{S}_t=s\tilde{S}_t dB_t$ 可以归结为对于 $\log\tilde{S}_t$ 用 Ito 公式．

2． 对于风险中性随机利率按方程 $dr_t=s\,dB_t+(a-br_t)dt$ 发展的金融债券, 在时刻 T 的单位币值在时刻 $t(\le T)$ 的折现为 $V(t,T)\overset{\Delta}{=}E(e^{-\int_t^T r_s ds}\mid r_s:s\le t)$．证明 $\tilde{V}(t,T)\overset{\Delta}{=}V(t,T)e^{-\int_0^t r_s ds}$ 是 (r_t) 鞅, 再证明存在函数 $F(t,x;T)$，使 $V(t,T)=F(t,r_t;T)$．

3．若上题中的 $F(t,x)$ 分别对 t,x 为一阶与二阶连续可微，证明它满足

$$F_t+\frac{s^2}{2}F_{xx}+a(b-x)F_x-xF=0，$$

$$F(T,x:T)=1．$$

再令

$$F(t,x;T)=e^{A(t,T)-xB(t,T)}，$$

并待定地确定

$$B(t,T)=\frac{1-e^{-a(T-t)}}{a},A(t,T)=(b-\frac{s^2}{2a^2})\{t-T+B(t)-\frac{s^2}{4a}B^2(t)\}．$$

4．对于在时刻 $L(L\le T)$ 到期的，执行价格为 K 的欧式利率看涨期权 $(F(L,r_L;T)-K)^+$，在时刻 $t(\le L)$ 的折现为 $C(t)\overset{\Delta}{=}E(e^{-\int_t^L r_s ds}(F(L,r_L;T)-K)^+\mid r_s:s\le t)$．证明 $\tilde{C}(t)\overset{\Delta}{=}C(t)e^{-\int_0^t r_s ds}$ 是 (r_t) 鞅．证明存在函数 $H(t,x;L,T)$，使 $C(t)=H(t,r_t;L,T)$．假定 $H(t,x;L,T)$ 分别对 t,x 为一阶与二阶连续可微，则它满足

$$H_t+\frac{s^2}{2}H_{xx}+a(b-x)H_x-xH=0，$$

$$H(L,x;L,T)=(F(L,x;T)-K)^+．$$

5．记 $L(t,x)=E(e^{lr_t-\int_0^t r_u du}\mid r_0=x)$．假定它分别对 t,x 为一阶与二阶连续可微，利用鞅 $M_t\overset{\Delta}{=}E(e^{lr_L-\int_0^L r_u du}\mid r_u:u\le t)$ 得到

$$L_t=\frac{s^2}{2}L_{xx}+a(b-x)L_x-xL．\ L(0,x)=e^{lx}$$

(可以得到这个方程的解

$$H(t,x;L,T)=F(t,x;T)\Phi(h+\frac{s_0}{2})-KF(t,x;L)\Phi(h-\frac{s_0}{2})，$$

其中

$$s_0=sB(L,T)\sqrt{\frac{1-e^{-2a(L-t)}}{2a}}，\quad h=\frac{1}{s_0}\log\frac{F(t,x;T)}{KF(t,x;L)})．$$

6． 对于随机利率按方程 $dr_t = s\,dB_t + (a - br_t)dt$ 发展的金融债券，在时刻T 的单位币值在时刻 $t(\le T)$ 的折现为 $V(t,T) \overset{\Delta}{=} E(e^{-\int_t^T r_s ds} \mid r_s : s \le t)$. 证明$\tilde{V}(t,T) \overset{\Delta}{=} V(t,T)e^{-\int_0^t r_s ds}$ 是(r_t)鞅.

7． 对于随机微分方程 $dr_t = s\sqrt{r_t}\,dB_t + (a - br_t)dt$. 令$G(t,x) \overset{\Delta}{=} E_x(e^{-lr_t - m\int_0^t r_u du})$， 假定它分别对$t, x$为一阶与二阶连续可微，利用$M_t \overset{\Delta}{=} E(e^{-lr_T - m\int_0^T r_u du} \mid r_u : u \le t)$是鞅，证明

$$G_t = \frac{s^2}{2} xG_{xx} + a(b - x)G_x - mxG = 0\text{，}\ G(0,x) = e^{-lx}\ .$$

龚光鲁，钱敏平著　应用随机过程教程 – 与在算法和智能计算中的应用
清华大学出版社，2003

第 14 章 在精算与风险模型中的应用

1 基本概念

１．１　保险中的利率概念

定义１４．１　（利率，名义利率与连续利率）

设实际年利率为 r_1，则折现系数定义为 $v \overset{\Delta}{=} \frac{1}{1+r_1}$．把与年利率 r_1 相等价的一年计息 m 次的名义利率 (nominal interest rate)记为 r_m，则它满足

$$(1+\frac{r_m}{m})^m = 1+r_1. \qquad (14.\ 1)$$

而

$$r = \lim_{m\to\infty} r_m$$

则就是第 13 章中的无风险银行利率，称为连续利率，在保险学中则称为利息强度(force of interest).

定义１４.２　（贴现率与名义贴现率）

对于一年一次计息的利率，在年终计算利息时就应该用年利率 r_1．但是如果在年初预付利息，则就要用贴现率，即预付利率 d，它就是利率的贴现率，即

$$d = \frac{r_1}{1+r_1}. \qquad (14.\ 2)$$

这个公式等价于

$$1+d+d^2+\cdots = 1+r_1, \qquad (14.3)$$

即：本利和＝1(元本金) + 预付率 + 预付率的预付率 + 预付率的预付率的预付率 $+\cdots$.

对于与年利率 r_1 等价的一年 m 次计息的名义利率 r_m，其相应的名义贴现率(预付利率) d_m 同样满足

$$1+\frac{d_m}{m}+(\frac{d_m}{m})^2+\cdots = 1+\frac{r_m}{m}. \qquad (14.\ 3)'$$

由此也可得名义贴现率的公式

$$d_m = \frac{r_m}{m+r_m}. \qquad (14.\ 2)'$$

１．２　生存模型的寿命分布与精算模型中的余寿

考察失效可有两种不同的角度．研究部件的失效时间和生物体的死亡时间，随着目的的不同，考虑的方式也会不同．例如，要知道某种药物对于动物死亡的影响时，我们往往不去追究被试验的动物的年龄．然而在精算中我们又必须考虑的年龄的差别．在精算界，人们把后一类归入所谓”有选择模型”，其意思是其对象是有选择的，例如，要考虑年龄的差别．

定义１４．３　（寿命分布，生存概率与失效率）

寿命是一个取非负值的随机变量 X，记其分布密度为 $p(t)$，分布函数为 $F(t)$，那么存活到时刻 t 生存函数(生存概率)为

$$S(t)\overset{\Delta}{=}1-F(t), \tag{14.4}$$

而时刻 t 的失效率（死亡率）定义为

$$l(t)\overset{\Delta}{=}\lim_{h\to 0} P(X\le t+h\mid X>t)=\frac{p(t)}{\int_t^{\infty}p(u)du}. \tag{14.5}$$

显见我们有

$$S(t)=e^{-\int_0^t l(u)du}. \tag{14.6}$$

典型的寿命分布，除常用的指数分布(由于它是无记忆的分布，在理论上不应直接应用于有选择模型)，对数正态分布外，还有

（1）Weibull 分布 $W(a,l)$　　$p(t)=l\cdot a\cdot t^{a-1}e^{-l\cdot t^a}I_{[0,\infty)}(t),(a,l>0)$，

其中 a 为形状参数，数学期望为 $l^{-\frac{1}{a}}\Gamma(1+\frac{1}{a})$，方差为 $l^{-\frac{2}{a}}[\Gamma(1+\frac{2}{a})-(\Gamma(1+\frac{1}{a}))^2]$． 又若 $x\sim \exp_l$，　则有 $h=x^{\frac{1}{a}}\sim W(l,,a)$．

（2）广义 Gamma 分布　　$p(t)=\frac{b}{\Gamma(g)}s^{-b\cdot g}t^{-b\cdot g-1}e^{-(\frac{t}{s})b}I_{[0,\infty)}(t)$.

（3）半边正态分布　　$F(t)=2(\Phi(t)-\frac{1}{2})I_{[0,\infty)}(t)$，其中 $\Phi(t)$ 为 $N(0,1)$ 的分布函数．

寿命的对数的分布还有

（1）极值分布　　$F(t)=1-e^{-e^{\frac{t-m}{s}}}$

（2）Logistic 分布　　$F(t)=\frac{1}{1+e^{-\frac{t-m}{s}}}$， 数学期望 m，方差 $\frac{1}{3}s^2p^2$．

其它被用作寿命分布的还有逆 Gauss 分布等．

精算中的死亡力度

如前所述，在精算中应该把投保人的年龄作为重要因素考虑．设年龄 x 的人的余寿 T 具有分布密度 $p_x(t)$，其分布函数为 $F_x(t)$． 那么，他在年龄 $x+t$ 时的危险率（死亡率，相

当于 ”顾及年龄” 的部件的失效率）应理解成他在时刻t死亡的密度：

$$l_x(t)=\lim_{\Delta t\to 0} P(T\le t+\Delta t \mid T>t)=\frac{p_x(t)}{1-F_x(t)}. \qquad (14.5)'$$

同样，生存时间超过t的概率为 $S_x(t)=e^{-\int_0^t l_x(u)du}$， 于是 $F_x(t)=1-e^{-\int_0^t l_x(u)du}$.

记号１４．４ 在国际精算界 $l_x(t)$ 被称为死亡力度．在概率界常用记号 $l_x(t)$，$F_x(t)$ 和 $S_x(t)$ 表示死亡力度，余寿分布和生存概率． 而在精算界，则有他们专用的传统记号，分别用 m_{x+t}， ${}_tq_x$，和 ${}_tp_x$．

典型的死亡力度模型有

（1） Gompertz 的指数死亡力度

$$l_x(t)=Ce^{x+t}.$$

（2） Makeham 死亡力度

$$l_x(t)=C_1+C_2e^{x+t}.$$

（3） Weibull 的幂死亡力度（余寿遵从 Weibull 分布）：

$$l_x(t)=C(x+t)^g,(g>0).$$

（4）线性指数死亡力度 （余寿分布称为线性指数分布，或 Rayleigh 分布）

$$l_x(t)=l+C(x+t).$$

(5) 阶梯形死亡力度．

(6) 盆状死亡力度，如

$$l_x(t)=\frac{b}{x+t+g}+d\cdot t\text{，或}\quad l_x(t)=\frac{b}{a}\left(\frac{x+t}{a}\right)^{b-1}e^{\left(\frac{x+t}{a}\right)^b}.$$

(7) 广义 Pareto 死亡力度

$$l_x(t)=a+\frac{b}{x+t+g}$$

等等．以上几个分布除线性指数分布和 Weibull 分布外，即使数学期望的解析式都很难求．一般需要用数值近似计算．

概念１２．５（生命表） 年龄x的人在当年内死亡的概率用表列出，称为生命表．它给出了在t为整数(以年为单位)时的死亡概率，在实际制定时，这些概率都是用统计频率近似得到的．而在t为非整数时，则可以用此生命表内插． 生命表是人寿保险投保操作的基本依据．

2 风险模型与破产理论介绍

２．１ 盈余过程与永不破产的概率

定义１４.５（理赔次数为 Poisson 过时的盈余过程） 设N_t是 Poisson 过程，它表示相继的理赔时刻，而各次理赔的金额是与理赔发生相互独立的独立同分布随机变量列$\{X_n\}$，其中 $X_n(>0)$ 表示第n次理赔金额．把时刻t前的累计索赔额记为S_t，即

$$S_t = X_1 + \cdots + X_{N_t}. \qquad (14.7)$$

它是强度 l 的复合 Poisson 过程．假定单位时间的投保费为c，而承担此项保险的公司的初始保证金(准备金)为x_0．那么在时刻t公司在此项保险上的盈余为

$$U_t = x_0 + ct - S_t. \qquad (14.8)$$

它是一个随机过程，称为盈余过程.

定义１４.６ 令$p = EX_1$．为了保证运行，保险公司必须要求 $c > lp$．记

$$\Lambda \overset{\Delta}{=} \frac{c}{lp} - 1, \qquad (14.9)$$

它称为相对安全负荷．

定义１４.７（破产时刻，破产赤字，最终破产概率，永不破产的概率）

随机时刻

$$T = \inf\{ t : U_t \le 0\} \qquad (14.10)$$

称为破产时刻，$|U_T|$称为破产赤字，它显然满足$U_{T-} > 0, U_T \le 0$．而

$$y(x_0) = P(T < \infty \mid U_0 = x_0) \qquad (14.11)$$

则是在$U_0 = x_0$的条件下最终破产的概率，则简称为破产概率． 类似地

$$y(x_0, t) = P(T \le t \mid U_0 = x_0) \qquad (14.12)$$

称为t前破产的概率． 记永不破产的概率为$R(x_0)$，则

$$R(x_0) = 1 - y(x_0) = P(U_t \ge 0, \forall t) = P(S_t - ct \le x_0, \forall t)$$

$$= P(\sup_{t \ge 0}(S_t - ct) \le x_0) \overset{\Delta}{=} P(L \le x_0), \qquad (14.13)$$

其中

$$L \overset{\Delta}{=} \sup_{t \ge 0}(S_t - ct)$$

是保险公司的最大损失． 可见永不破产的概率$R(x_0)$正是最大损失的分布函数．类似地还有t前不破产的概率

$$R(x_0,t)\overset{\Delta}{=}1-y\,(x_0,t)\,.\qquad (14.\ 14)$$

显见有　$R(x_0,\infty)=R(x_0)$.

在本书中，如果不作特别声明，恒假设理赔额 X_i 是有分布密度 $p_X(x)$ 的随机变量．于是累计理赔额 S_t （它是复合 Poisson 过程）的分布函数为 $F_S(x,t)$ 具有密度函数，记为 $p_S(x,t)$．从而盈余过程 U_t 的分布函数为

$$F_U(x,t)=P(x_0+ct-S_t\le x)=1-F_S(x_0+ct-x,t)\,,\qquad (14.15)$$

并有密度，记为 $p_U(x,t)$，表示准备金为 x_0 时，盈余过程在时刻 t 的分布密度．显见有

$$p_U(x,t)=p_S(x_0+ct-x,t)\,.\qquad (14.\ 15)'$$

２．２　t 前不破产的概率的公式与估计

准备知识

定义１４．７　随机变量序列 $\{X_n\}$ 称为可交换的随机序列，如果对于任意 m 及 $\{1,2,\cdots,m\}$ 的任意一个排序 $\{i_1,i_2,\cdots,i_m\}$，均有 $\{X_{i_1},X_{i_2},\cdots,X_{i_m}\}$ 与 $\{X_1,X_2,\cdots,X_m\}$ 同分布．可交换的随机序列 X_n 的部分和 $S_n\overset{\Delta}{=}x+X_1+\cdots+X_n,(S_0=0)$ 称为具可交换增量的随机序列．

例１４．８　独立同分布随机变量的部分和是最简单的具可交换增量的随机序列．

例１４．９　设 N_t 是以 $\{t_n\}$ 为更新流的更新过程，而其独立同分布的更新间隔为 $T_1,\cdots,T_n,\cdots$．这时，在条件 $N_t=n$ 下，$\{t_m\}_{m\le n}$ 对于条件概率 $P(*\,|\,N_t=n)$ 而言，是具有可交换增量的随机序列

与随机徘徊相类似，具可交换增量的随机序列也有对称原理，它是随机徘徊的对称原理的推广．

命题１４．１０　（可交换增量的随机序列对称原理）设 S_n 为具可交换增量的随机序列，且 $S_0=0$，则有

$$P(S_i>0,(i<n),S_n\in[a,b])=P(S_i<S_n,(i<n),S_n\in[a,b])\,.\qquad (14.\ 16)$$

(其证明几乎可以照搬第 3 章中随机徘徊的对称原理(定理３．３０)的证明)

命题１４．１１　(Dwass –Dinges 定理)

设 S_n 为具可交换增量的随机序列 $S_n=x+X_1+\cdots+X_n$，其中 X_i 取值于 $\{1,0,-1,-2,\cdots\}$．$S_0=x$(整数)．那么，在 $m>x$ 时有

$$P(S_i < m, (i < n), S_n = m) = \frac{m-x}{n} P(S_n = m). \quad (14.\ 17)$$

(**证明几乎可以照搬第** 3 **章定理３．３４的证明**).

[**注**] **可交换的随机序列可以由条件独立同分布随机变量列来描述. 这就是著名的** de Finetti **定理. 我们把它放在下面以供参考**

de Finetti 定理 若$\{X_n\}_{1\le n<\infty}$是可交换的随机序列 (只有有限个随机变量时定理不真), 则存在一个随机变量h, 使在h已知的条件下, $\{X_n\}$为条件独立同分布的. 即对于任意$m, x_1, \cdots, x_n, y$, 恒有

$$P(X_1 \le x_1, \cdots, X_m \le x_m \mid h = y) = \prod_{k=1}^{m} P(X_k \le x_k \mid h = y).$$

１．无准备金情形的永不破产概率

引理１４．１２ 若$U_0 = x_0 = 0$, 则对$x > 0$, 我们有

$$P(U_s \ge 0, x < U_t \le x + dx, \forall s < t) = P(U_s < U_t, x < U_t \le x + dx, \forall s < t)$$

$$= \frac{x}{ct} P(x < U_t \le x + dx). \quad (14.\ 18)$$

(**证明大意** 由于盈余过程U_t是由复合 Poisson 过程构成的, 所以它是独立增量过程. 于是它在相等时间间隔的采样是独立随机变量的和, 因而采样列是具有可交换增量的随机变量列. 对它应用对称原理, 再让采样间隔趋于 0, 便得第一个等式. 再对$\frac{1}{c}U_t$用 Dwass-Dinges 定理, 并让采样间隔趋于 0, 便得第二个等式).

(1) **无准备金时t前不破产概率$R(0,t)$的第一个计算公式**

因为$x_0 = 0$, **所以**$U_t \le ct$. **在**(14. 18)**两边对**x**从** 0 **到**ct**积分, 我们得到**

$$R(0,t) = P(U_s \ge 0, \forall s \le t) = P(U_s \ge 0, U_t \le ct, \forall s \le t) = \frac{1}{ct}\int_0^{ct} x p_U(x,t)dx.$$

对于上式右方用分部积分, 再用变量替换$ct - x = y$, **并用**(14.15), **则**

$$\text{上式右方} = \frac{1}{ct}[ctF_U(ct,t) - \int_0^{ct} F_U(x,t)dx]$$

$$= \frac{1}{ct}[ct - \int_0^{ct}(1 - F_S(ct - x, t))dx] = \frac{1}{ct}\int_0^{ct} F_S(y,t)dy.$$

于是我们得到如下的第一个公式

$$R(0,t) = \frac{1}{ct}\int_0^{ct} F_S(y,t)dy. \quad (14.\ 19)$$

[**注**] **此公式在理赔额无分布密度时仍然正确, 证明只需作一些必要的修改.**

(2). **无准备金时t前不破产概率$R(0,t)$的第二个计算公式**

由前面的推导过程可以看出

$$R(0,t)=\frac{1}{ct}[ct-\int_0^{ct}(1-F_S(y,t))dy]=\frac{1}{ct}[ct+\int_{ct}^{\infty}(1-F_S(y,t))dy-\int_0^{\infty}P(S_t>y)dy].$$

再利用 $\int_0^{\infty}P(S_t>y)dy=ES_t=ltp$**，并用相对安全负荷** $\Lambda=\frac{c}{lp}-1$ **作为参数，便得另一个表达式**

$$R(0,t)=\frac{\Lambda}{1+\Lambda}+\frac{1}{ct}\int_{ct}^{\infty}P(S_t>y)dy. \qquad (14.20)$$

我们把它们综合为下面的定理。

定理１４．１３ 对无准备金且以复合 Poisson **过程理赔的风险模型，在时刻** t **以前破产的概率为**

$$R(0,t)=\frac{1}{ct}\int_0^{ct}F_S(y,t)dy=\frac{\Lambda}{1+\Lambda}+\frac{1}{ct}\int_{ct}^{\infty}P(S_t>y)dy.$$

(3) **无准备金时前不破产概率** $R(0,t)$ **的上界估计**

假定累次理赔额 X_i **的二阶矩有限．由** Chebyshev **不等式，我们有**

$$P(S_t>y)=P(S_t-ES_t>y-ltp)\le\frac{Var(S_t)}{(y-ltp)^2}=\frac{ltEX_1^2}{(y-ltp)^2}.$$

于是由(14. 20)**得到上界估计**

$$R(0,t)\le\frac{\Lambda}{1+\Lambda}+\frac{lEX_1^2}{c}\int_{ct}^{\infty}\frac{1}{(y-ltp)^2}dy\le\frac{\Lambda}{1+\Lambda}+\frac{lEX_1^2}{ct}\frac{1}{c-lp}. \qquad (14.21)$$

(4). **无准备金时永不破产概率** $R(0)$ **的公式**

$$R(0)=R(0,\infty)=\frac{\Lambda}{1+\Lambda} \qquad (14.22)$$

可见最终破产的概率为

$$y(0)=\frac{1}{1+\Lambda}, \qquad (14.23)$$

即相对安全负荷 $\Lambda=\frac{c}{lp}-1$ **越大，则最终破产概率越小，这与直觉并无二致．**

２．有准备金情形下的不破产概率

现在设 $U_0=x_0>0$**．我们可以把有准备金情形化为无准备金情形．显见**

$$R(x_0,t)=P(U_s\ge0,(\forall s\le t))=P(U_t\ge0)-P(U_t\ge0,\exists s<t\text{使}U_s=0)$$

$$=P(S_t\le ct+x_0)-$$

$$-\int_0^t P(U_t\ge0,\exists s\in[p,p+dp)\text{使}U_s=0,\text{且}\forall r\in(s,t)\text{有}U_r>0)dp. \qquad (14.23)$$

由于 S_t **在** $[p,p+dp)$ **中不变化的概率为** $1-o(dp)$**，从而** $P(\exists s\in[p,p+dp)\text{使}U_s=0)$ **与**

$P(U_p \in (-cdp, 0])$ 只差 $o(dp)$ (它们都近似地等于在长为 dp 的时间段只有一次理赔的概率). 再则，由 U_t 的独立增量性可知它也是 Markov 过程, 直观地利用全概率公式 (如果要追究数学的严格性, 就需要注意破产时刻 T 是随机变量. 在把 T 作为 "现在"时刻, 利用 U_t 是独立增量过程, 仍然可以证明它对于这种随机的 "现在" 仍具有 Markov 性 {对随机的 "现在" 仍有 Markov 性的随机过程, 称为具有强 **Markov** 性． 本书中略去这较多测度论知识的推导}, 注意到此时 $dU_p = cdp$, 于是(14. 23)的右方为

$$= F_S(x_0 + ct, t) - \int_0^t p_U(0, p) cdp R(0, t-p) .$$

再用(14. 15), 就得到把有准备金情形化为无准备金情形的 t 前不破产概率的公式，即下面的定理.

定理１４.１４

$$R(x_0, t) = F_S(x_0 + ct, t) - c\int_0^t R(0, t-p) p_S(x_0 + cp, p) dp . \quad (14\ 24)$$

[注] 此公式在随机理赔额无分布密度时只需作相应的修改．

２．３　有准备金时最终破产概率的上界与调节系数

定理１４．１５　下述关于 s 的方程

$$l + sc = l\int_0^\infty e^{sx} p_X(x) dx \qquad (14.\ 25)$$

存在唯一正解 R (即 R 满足 $1 + pR(1+\Lambda) = \int_0^\infty e^{Rx} p_X(x) dx$). 它给出了最终破产概率 $y(x_0)$ ($-\infty < x_0 < \infty$, 即容许初始盈余为负—-负债的情形) 随初始盈余衰减的指数界：

$$y(x_0) \le e^{-Rx_0} . \qquad (14.\ 26)$$

这个 R 称为调节系数．

[注] 如果将理赔额的分布密度的 Laplace 变换记为 $L_{P_X}(s) \overset{\Delta}{=} \int_0^\infty e^{-sx} p_X(x) dx$, 那么 R 为非线性方程

$$l[L_{p_X}(-s) - 1] = sc \qquad (14．26)'$$

的正根, 它可以用数值分析方法求得.

证明　首先，0 显然是此方程的一个解． 再则, 方程(14.25)的左方是 s 的线性函数, 右方则是 s 的趋于 ∞ 的凸函数， 而在 $s = 0$ 处左边的导数大于右边的导数, 因此, 此方程还存在一个正解．我们将第 n 次理赔前破产的概率记为 $y_n(x_0)$ ．那么, $y_n(x_0) \le y(x_0)$ 且 $y_n(x_0) \to y(x_0)$ ．于是定理的证明只需对 n 归纳地证明

$$y_n(x_0) \le e^{-Rx_0} . \qquad (14.\ 27)$$

注意 $y_0(x_0) = I_{(-\infty, 0]}(x_0)$, 即 $n = 0$ 时(14. 27)自然成立． 今作归纳法假设 $n-1$ 时(14. 27)

正确．对于n的情形，利用第一次理赔时刻服从指数分布，在计算概率$y_n(x)$时，对第一次理赔时刻与理赔额运用全概率公式，再用归纳法假定及调节系数的定义，我们得到

$$y_n(x_0)=\int_0^\infty l e^{-lt}P(n\text{次理赔前破产}\,|\,\text{首次理赔时刻为}t,\text{理赔额为}y)p_X(y)dydt$$

$$=\int_0^\infty le^{-lt}dt\int_0^\infty y_{n-1}(x_0+ct-y)p_X(y)dy\le\int_0^\infty le^{-lt}dt\int_0^\infty e^{-R[x_0+ct-y]}p_X(y)dy$$

$$=\frac{l}{l+Rc}e^{-Rx_0}\int e^{Ry}p_X(y)dy=e^{-Rx_0},$$

其中最后一个等号得自R的定义．

[注 1] 递推公式

$$y_n(x_0)=\int_0^\infty le^{-lt}dt\int_0^\infty y_{n-1}(x_0+ct-y)p_X(y)dy \qquad (14.\ 28)$$

正给出了最终破产概率的递推近似.

[注 2] (一般的理赔更新流情形) 如果理赔流不是指数流，而是一般的更新流，那么调节系数 R 定义应为下述方程的解s：

$$Ee^{s(X_1-cT_1)}=1, \qquad (14.\ 29)$$

其中T_1为首次理赔时刻．同样可以证明 $y(x_0)\le e^{-Rx_0}$． 注意当理赔流为指数流时，调节系数用(14. 29)定义与用((14. 25)定义是一致的．

定理１４．１５ (调节系数的估计)

$$R<2\Lambda\frac{EX_1}{EX_1^{\ 2}}. \qquad (14.\ 30)$$

又若理赔额X_i是有界随机变量，且其上界为M，则还有下界估计

$$R>\frac{1}{M}\log(1+\Lambda). \qquad (14.\ 31)$$

证明　由R满足

$$1+pR(1+\Lambda)=\int_0^\infty e^{Rx}p_X(x)dx>\int_0^\infty(1+Rx+\frac{(Rx)^2}{2})p_X(x)dx=1+REX_1+\frac{1}{2}R^2EX_1^{\ 2}$$

立得(14.30)．另一方面，在区间$[0,M]$的两端，凸函数e^{Rx}与线性函数$\frac{x}{M}(e^{RM}-1)+1$相等，由此推出

$$e^{Rx}\le\frac{x}{M}(e^{RM}-1)+1.$$

于是

$$1+pR(1+\Lambda)=\int_0^\infty e^{Rx}p_X(x)dx\le\int_0^\infty(1+\frac{x}{M}(e^{RM}-1))p_X(x)dx=1+\frac{(e^{RM}-1)}{M}p.$$

因此　$1+\Lambda\le\frac{e^{RM}-1}{RM}<e^{MR}$，便得(14.31)．

［注］理赔额存在密度 $p_X(x)$ 的假定也是不必要的，其证明也只需要把普通积分改为 Stieltjes **积分**.

２．４　破产概率的方程

由于理赔额超过 x 的概率为 $1-F_X(x)$，对于 $x\geq 0$，在(14. 28)中，令 $n\to\infty$，便得到破产概率满足积分方程

$$y(x)=\int_0^{\infty} l e^{-lt}dt\int_0^{\infty} y(x+ct-y)p_X(y)dy. \qquad (14.\ 32)$$

令 $u=x+ct$，那么

$$y(x)=\int_x^{\infty} l e^{-\frac{l(u-x)}{c}}du\int_0^{\infty} y(u-y)p_X(y)dy,(x\geq 0). \qquad (14.\ 33)$$

另一方面，显见应有

$$y(x)=1,(x<0). \qquad (14.\ 34)$$

对(14. 33)求导数得到

$$y'(x)=\frac{l}{c}[y(x)-\int_0^{\infty} y(x-y)p_X(y)dy].$$

再利用(14. 34)就得到下面的定理.

定理１４．１６　最终破产概率满足的积分微分方程:

$$y'(x)=\frac{l}{c}(y(x)-\int_0^{x} y(u)p_X(x-u)du-[1-F_X(x)]),(x>0). \qquad (14.\ 35)$$

它也可简单地写成永不破产概率的积分微分方程:

$$R'(x)=\frac{l}{c}[R(x)-\int_0^{x} R(u)p_X(x-u)du]. \qquad (14.\ 35)'$$

例１５．１７　在理赔额服从指数分布及混合指数分布时，可以由(14. 35)再求导数，导出 y 满足的常系数常微分方程，求解便可得 y 的明显表达式．例如，在理赔额服从参数为 m 的指数分布时，调节系数为 $R=\frac{m\Lambda}{1+\Lambda}$，最终破产概率为 $y(x_0)=\frac{1}{1+\Lambda}e^{-Rx_0}$．

［注１］理赔次数可以不必局限为 Poisson 过程，一般地可以是一个更新过程，其中最简单的是 Erlang(2) 过程．相应地就有 Erlang(2) 盈余过程．

［注２］还有大数定律 $\frac{U_t}{N_t}\xrightarrow{t\to\infty}\frac{c-lEX_1}{l}$，从而有 $U(\infty)=\infty$．

［注３］一般重点关心的问题还有：在初始保证金为 x_0 时，集体风险理论中的破产时刻 T，破产前的盈余 U_{T-} 与破产时刻的赤字 $-U_T$．这三个随机变量组 $(T,U_{T-},-U_T)$ 如果存在联合密度

$f(t,x,y|x_0)$，那么它将提供重要的统计信息．若假定破产前的盈余为 x 而破产时的赤字为 $y(<0)$ 时的惩罚权重函数为 $w(x,y)$．那么平均折现为

$$f(x_0)=\int_{-\infty}^{0}\int_{0}^{\infty}\int_{0}^{a} e^{-rt} f(t,x,y|x_0)dtdxdy,$$

其中 r 为利率．

在实用中常把时间取为离散采样．这时有关的一些概率量，常可写成迭代形式（对应于连续时间情形的更新方程），这时非常有利于作数值计算．

2．5　保险费的效用函数与保险费策略的制定

如果抽象地考虑风险，保险公司承担的风险 h 是一个随机变量（例如，h = 复合 Poisson 过程 S_t 的情形）．公司为此收取保险费是依赖于此风险的一个正数，记成 $P(h)$．一次投保使保险公司有随机收入 $P(h)-h$．比率 $\frac{P(h)}{Eh}$ 称为保费率．在无竞争条件下，在制定保险费策略 $P(\cdot)$ 时，通常要求满足：

(1)　$Eh \le P(h) \le$ 某个上界；

(2)　$P(\cdot)$ 应像数学期望，但是一般没或有可加性，而只对多个独立的风险有可加性；

(3)　对于两个风险 h,V，与条件期望类似地，可以定义条件风险 $P(h|V)$，它是一个依赖于 V 的随机变量，因而条件风险也可以再投保，而且应该要求保费策略满足 $P(P(h|V))=P(h)$．

最平凡的取法是令 $P(h)=Eh$，称为净保策略． 但是这不容易被保险公司所接受．而通常的其它情形有　$P(h)=(1+a)Eh$，$P(h)=Eh+aVar(h)$，$P(h)=Eh+a\sqrt{Var(h)}$ 等等，　但是它们都并不能保证 (1) — (3) 全部满足． 所以人们提出了效用函数和非线性数学期望等概念．

保险公司获得的盈余 x 的实际作用，可以用一个函数 $u(x)$ 来公平地表达，它应该满足 $u'(x)>0, u''(x)\le 0$，即它是一个可微的严格的递增凹函数，称为效用函数．

有一种考虑，是使用效用函数 $u_a(x)=\frac{1-e^{-ax}}{a}$（显见它关于 a 是递增的）作用下的平均原理，即对于固定的 x，将 h 作为变量，在方程

$$u_a(x)=E[u_a(x+P-h)] \qquad (14.\ 36)$$

中解出一个与初始盈余 x 无关的 P 作为 h 的函数，取之为 $P(h)$，并将它记为 $P_a(h)$．这时有（请读者自行验证）

$$P_a(h)=\frac{1}{a}\log Ee^{ah}, \tag{14.37}$$

再定义

$$P_a(h\,|V)=\frac{1}{a}\log E(e^{ah}\,|V).$$

这种计算保险费用的策略，称为指数零效用原理．它满足以上（1）－（3）的所有要求（请读者自行验证）．在风险 h 为累计理赔额为复合 Poisson 过程 S_t 时，指数零效用原理就是说，盈余 $x+P(S_t)-S_t$ 的效用 $u(x+P(S_t)-S_t)$ 是一个鞅，即效用是"公平地"发展的，这样的原则容易得到保方与被保方认可．

当然，对于一般的效用函数，也可以类似地用平均原理给出收取保险费的策略.

例１４．１７　设保险的风险 h 为累计理赔额为复合 Poisson 过程 S_t．假定按指数零效用原理收取保险费 $P_a(t)\stackrel{\Delta}{=}P_a(S_t)=\frac{1}{a}\log Ee^{aS_t}$，那么单位时间的投保费为

$$c=\frac{P_a(t)}{t}=\frac{1}{at}\log Ee^{aS_t}=\frac{1}{a}Ee^{aX_1}\quad\text{(读者自行验证)},$$

其中 X_1 为一次理赔额度．此时的调节系数 $R=a$（读者自行验证）．

非线性条件期望是由彭实戈提出的，它由向后随机微分方程的解给出．至今还未在保险领域起用．我们认为这种概念将能在制定保费策略上发挥重要的理论作用．

２．６　最大损失的分布

在复合 Poisson 过程作为累计理赔时，我们求最大损失 $L=\max_{t>0}\{S_t-ct\}$ 的分布．为此，只需求它的 Laplace 变换．由(14.13)，最大损失 L 的分布函数就是永不破产概率 $R(x)$，$P(L=0)=R(0)=\frac{\Lambda}{1+\Lambda}$．用分部积分便得

$$Ee^{-Ls}=R(0)+\int_{(0,\infty)}e^{-sx}dR(x)=R(0)-\int_0^\infty e^{-sx}y'(x)dx=1-sL_y(s), \tag{14.38}$$

此处

$$L_y(s)=\int_0^\infty e^{-sx}y(x)dx.$$

而对于理赔额 X_1 的密度 $p_X(x)$ 的 Laplace 变换，用分部积分有

$$L_{p_X}(s)\stackrel{\Delta}{=}\int_0^\infty e^{-sx}p_X(x)dx=1-s\int_0^\infty e^{-sx}(1-F_X(x))dx.$$

另一方面，由(14.35)求 Laplace 变换得到(注意卷积的 Laplace 变换是 Laplace 变换的乘积)

$$\frac{c}{l}\int_0^\infty e^{-su}y'(u)du=\int_0^\infty e^{-su}y(u)du-\int_0^\infty e^{-su}y(u)du\int_0^\infty e^{-su}f_X(u)du-\int_0^\infty e^{-su}(1-F_X(u))du$$

即

$$\frac{c}{l}[R(0)-1+sL_y(s)]=L_y(s)[1-L_{p_X}(s)]-\frac{1}{s}[1-L_{p_X}(s)].$$

解此方程得

$$L_y(s)=\frac{\frac{c}{l}[1-R(0)]-\frac{1}{s}[1-L_{p_X}(s)]}{\frac{cs}{l}-[1-L_{p_X}(s)]}.$$

再用（１４．３８），便得到最大损失 L 的分布的 Laplace 变换的明显表示式如下

$$Ee^{-sL}=1-sL_y(s)=\frac{\frac{cs}{l}R(0)}{\frac{cs}{l}-[1-L_{p_X}(s)]}.$$

由它在 $s=0$ 的展开式易得 $EL=\frac{EX_1{}^2}{2\Lambda EX_1}$， $EL^2=\frac{EX_1{}^3}{3\Lambda EX_1}+\frac{1}{2}\left(\frac{EX_1{}^2}{\Lambda EX_1}\right)^2$.

[注] 第 2 节的部分内容选材自

汉斯 U. 盖伯，数学风险论导引 (成世学，严颖 译)，世界图书出版公司，1997

3 考虑利率与投资的保险模型简述

设(连续)利率为 r．假定收取的保险金存入银行． 那么盈余过程应满足

$$dU_t=cdt+rU_tdt-dS_t.$$

此处由于 S_t 是强度为 l 的复合 Poisson 过程，因而是独立增量过程． 可以与 Ito 积分类似地定义函数对于 S_t 的积分，且更为简单， 即在作积分和时，不会如 Ito 积分那样依赖于取左端点还是取右端点． 对此方程进行积分便得盈余过程的表达式

$$U_t=ue^{rt}+c\frac{e^{rt}-1}{r}-\int_0^t e^{r(t-p)}dS_p.$$

还可以考虑如下类型的随机利率

$$r(t)=r_0+\Lambda(t)-l't,$$

其中 $\Lambda(t)$ 是一个强度为 l' 的 Poisson 过程． 此时有单位币值 $L(t)=e^{\Lambda(t)+(r-l)t}$ ．也可考虑利率 $r=r_0+sB_t$ 的情形．最后还可以考虑用保险金投入风险市场购买投资组合的情形． 可参考

Dickson, Math & Economics 11,191-207,1992.

Dufresne & Gerber, ibid, 7,193-199,1988.

Egidio Dos Reis, ibid, 12,23-38,1993

Geber, H. U. and Shiu, E.S.W., On the them value of ruin, North Amer. Acturial J., 2(1), 48-78．

习题 14

1. 求 $l \to \infty$ 时，累计理赔的近似分布．

2. 只考虑整数理赔时刻，并设它们以二值分布独立同分布地到达．于是累计理赔 S_{N_n} 为复合负二项随机变量列．试讨论相应的破产理论．

3. 设风险 h 有分布密度 $p(x)$，$M(s)=Ee^{sh}$ 是 h 的矩母函数．定义风险 h 的 Esscher 变换为分布密度为 $\dfrac{e^{ax}p(x)}{M(a)}$ 的随机变量 $h^{(a)}$．证明

$$(a^2(\frac{\log M(a)}{a})')'=a[\frac{M''(a)}{M(a)}-(\frac{M'(a)}{M(a)})^2]=Var(h^{(a)})．$$

4. 验证 (14. 37)．

5. 验证对于复合 Poisson 为累计理赔时，由效用函数 $u_a(x)=\dfrac{1-e^{-ax}}{a}$ 按平均原理制定的单位时间的保费为 $\dfrac{P_a(t)}{t}=\dfrac{1}{a}Ee^{aX_1}$，而调节系数为 a．

6 当理赔流为指数流时，证明调节系数用(14. 29)定义与用((14. 25)定义是一致的．

龚光鲁，钱敏平著　应用随机过程教程 – 与在算法和智能计算中的应用
清华大学出版社，2003

第 15 章 与数据建模有关的几个算法

1　EM 算法 – 隐状态变量分布中参数的最大似然估计

1．1　 EM 算法的基本想法

在数据资料不全时，由已有的资料$\vec{Y}$估计缺失变量$\vec{X}$，或在观测到的资料$\vec{Y}$并不是状态变量时,估计状态变量$\vec{X}$时（或者估计其分布密度$f(J,\vec{x})$中的未知参数J)，就与古典统计很不相同，此时需要用观测到的资料$\vec{Y}$，同时估计$\vec{X}$与未知参数J．

这样的估计将面临如下困难： 如果把在参数J下的期望记为E_J，那么，在估计状态变量$\vec{X}$时，估值当然应该用条件期望$\hat{X}=E_J(\vec{X}\,|\,\vec{Y})$（ 如果在$\vec{Y}=\vec{y}$及$J$的条件下，$\vec{X}$的 Bayes 分布密度$f(\vec{x}\,|\,\vec{y},J)$为已知，则也常用 Bayes 估计$\hat{X}=\int\vec{x}f(\vec{x}\,|\,\vec{Y},J)d\vec{x}$）．然而这时就需要知道参数$J$的值；另一方面，为了知道$J$，又必须先知道$\vec{X}$的估值$\hat{X}$(作为状态的已知样本值)．这样，估计状态与估计未知参数之间是耦合的．

在统计中通常对付这类困难的解耦方法是：假定一个已知，迭代地分别交替估计它们中的另一个，直至稳定．此类算法通称为 EM 算法，较为确切的表达是：

(1) 设置初值 J_0；

(2)（E-步骤）对$n\geq 0$, 令 $\hat{X}^{(n)}=E_{J_n}(\vec{X}\,|\,\vec{Y})$ (或用 Bayes 估计 $\hat{X}^{(n)}=\int\vec{x}f(\vec{x}\,|\,\vec{Y},J_n)dx$)；

(3)（M -步骤）（修正J的估计） 取J_{n+1}使之满足：

$$\log f(J_{n+1},\hat{X}^{(n)})=Max_J\log f(J,\hat{X}^{(n)}),$$

其中 E-步骤为取条件期望（ Expectation ），而 M-步骤为取最大（ Maximum ）．这种交替迭代的方法， 称为简单的 EM 方法．

这个算法的构思很简单，但计算量过大，且一般很难看出是否稳定．为了克服这个缺点， Dempster, Laird 和 Rubin 提出了直接递推估计J的想法(仍旧称为 EM 方法)，这种经过本质改进后的方法，至少在直观上看起来有稳定趋势．

1．2　 Rubin 算法

假定$(\vec{X},\vec{Y})$具有联合分布密度$f(J,\vec{x},\vec{y})$

Rubin 算法的核心构思为： 直接使用状态变量$\vec{Y}$的分布密度$g(J,\vec{y})$代替$(\vec{X},\vec{Y})$的分

布密度 $f(J,\vec{x},\vec{y})$，来求 J 关于 $g(J,\vec{y})$ 的最大似然估计 $\hat{J}$． 也就是求使 $\log g(J,\vec{Y})$ 达到最大的 $\hat{J}$．

由于 $g(J,\vec{y})$ 在实际上并不好求，Dempster– Laird- Rubin 利用了下述等式

$$f(J,\vec{x},\vec{y})=f(J,\vec{x}\mid\vec{y})\ g(J,\vec{y}), \tag{15.1}$$

于是

$$\log g(j,\vec{Y})=\log f(j,\vec{X},\vec{Y})-\log f(j,\vec{X}\mid\vec{Y}), \tag{15.1'}$$

其右方是状态变量的对数似然密度与对数条件似然密度的差．对等式两边取关于 $\vec{Y}$ 的条件期望得到

$$L(j)\overset{\Delta}{=}\log g(j,\vec{Y})\ \ (=E_J(\log g(j,\vec{Y})\mid\vec{Y}))$$

$$=E_J([\log f(j,\vec{X},\vec{Y})]\mid\vec{Y})-E_J\log[f(j,\vec{X}\mid\vec{Y})]$$

$$\overset{\Delta}{=}Q(j\mid J)-H(j\mid J). \tag{15.2}$$

在观测资料是 $\vec{y}$ 的时候(即 $\vec{Y}=\vec{y}$ 的时候)，$Q(j\mid J)$ 的表达式为

$$Q(j\mid J)=\int\log f(j,\vec{x},\vec{y})f_{\bar{X}|\bar{Y}}(q,\vec{x}\mid\vec{y})d\vec{x},$$

而 $H(j\mid J)$ 的表达式为

$$H(j\mid J)=\int\log f(j,\vec{x},\vec{y})f_{\bar{X}|\bar{Y}}(q,\vec{x}\mid\vec{y})d\vec{x}$$

$$=\int\log f_{\bar{X}|\bar{Y}}(j,\vec{x}\mid\vec{y})f_{\bar{X}|\bar{Y}}(q,\vec{x}\mid\vec{y})d\vec{x}.$$

为了求 $L(J)$ 的极大值点，我们将沿用第 10 章中对得到隐 Markov 模型的模型参数估计的 Baum-Welsh 算法的思想．为此注意

$$L(j)-L(J)=[Q(j\mid J)-Q(J\mid J)]+[H(J\mid J)-H(j\mid J)]$$

$$=[Q(j\mid J)-Q(J\mid J)]+\int[\log\frac{f_{\bar{X}|\bar{Y}}(J,\vec{x}\mid\overrightarrow{Y})}{f_{\dot{X}|\dot{Y}}(j,\vec{x}\mid\dot{Y})}]f_{\dot{X}|\dot{Y}}(J,\vec{x}\mid\vec{Y})d\vec{x}$$

$$\geq[Q(j\mid J)-Q(J\mid J)], \tag{15.3}$$

其中用到第２项是相对熵的非负性．由(15．3)可知：

只要 $Q(j\mid J)-Q(J\mid J)\geq 0$ 就有 $L(j)-L(J)\geq 0$．

所以，为了将 J 的较粗估计 J_n，修改为较精的估计 J_{n+1}，只需找 J_{n+1} 使 $L(J_{n+1})-L(J_n)\geq 0$．也就是只需要找 J_{n+1} 使

$$Q(J_{n+1}\mid J_n)-Q(J_n\mid J_n)\geq 0\,. \qquad (15.4)$$

于是我们就得到如下的 Dempster– Laird- Rubin 的 EM 算法，简称 Rubin 算法或 EM 算法：

(1) 设置初值 J_0；

(2)（E-步骤）对于 $n\geq 0$，计算 $Q(j\mid J_n)=\int \log f(j,\vec{x},\vec{y})f_{\bar{X}|\bar{Y}}(q,\vec{x}\mid\vec{y})d\vec{x}$；

(3)（M -步骤）（修正 J 的估计） 取 J_{n+1} 使

$$Q(J_{n+1}\mid J_n)=Max_j\, Q(j\mid J_n)\,. \qquad (15.5)$$

由于将 J_n 修正为 J_{n+1} 时，$Q(J_{n+1}\mid J_n)$ 比 $Q(J_n\mid J_n)$ 大，所以 $L(J_n)$ 是不减的，这就说明 J_n 有收敛的趋势．Rubin 证明了，在一定的条件下，J_n 确实按概率收敛于某个 $\hat{J}$．于是我们可以合理地把 $\hat{J}$ 作为分布的参数 J 的较佳的估计

注 1　$L(J)$ 是多峰函数的时候，$\hat{J}$ 有可能是 $L(J)$ 的局部最大值，甚至是鞍点，为了纠正这种不足，一般可以从多个初始值出发，找到多个 $\hat{J}$ 值进行比较．

注 2 所谓“ 缺失资料 ”可见之于两种情形：一种是观测资料的丢失，另一种是人为地设置一些“后台操作的 ”辅助随机变量，称之为潜变量，将他们视为缺失的资料，即将他们视为不能直接测量到的状态随机变量．这种潜变量的取法可以非常灵活，依赖于人们对于问题的认识，经验的积累与对于技术掌握的成熟程度．从数学处理的角度考虑，最好能选取潜变量 $\vec{X}$，使它与观测随机变量 $\vec{Y}$ 的联合分布是指数型分布（参见第 1 章），这时 E-步骤就很容易计算．

注 3　在离散随机变量的情形，E－步骤中的积分应该相应的改为求和。

以上算法也同样将隐马氏模型中的 Baum-Welsh 算法，纳入了 EM 算法的框架．

１．３　EM 算法的变通 – 广义 EM 算法

在实际计算中，又因为 E-步骤的计算量很大，人们常常并不真去计算条件期望，而是采用随机模拟．即：用在 $(J_n,\vec{Y})$ 已知的条件下，用随机模拟得到 $\vec{X}$ 的 Bayes 分布的若干独立随机数，代入 $\log f(j,\vec{X})$ 后作样本平均，用来代替 $Q(j\mid J_n)$．这就大大地减少了计算量．实践表明这样的简化，常可以得到相当满意的效果．

同样，也因为 M-步骤的计算量也很大，人们也常常并不去算最大，而是任意找一个 j，只要满足 $Q(j\mid J_n)-Q(J_n\mid J_n)\geq 0$，就取 j 为 J_{n+1}．这样简化了的算法，称为广义

EM 算法，常称为 GEM 算法．

[注]　在应用 EM 方法时，为了避免 M-步骤中求最大值的复杂计算，还可以采取其它的灵活的替代方法．例如有

ECM 算法(Conditional Maximum, CM)，　在多个参数的时候，例如$J=(J_1,J_2)$的情形，如下的交替地求条件极值的最大化的方法，也常被用来代替 M-步骤：

(1)　先取$J_2^{(n+1)}$满足

$$Q((J_1^{(n)},J'^{(n+1)}_2)\mid J^{(n)})=\max_{J_2'}Q((J_1^{(n)},J'_2)\mid J^{(n)});$$

(2)　再取$J_1^{(n+1)}$满足

$$Q((J_1^{(n+1)},J_2^{(n+1)})\mid(J_1^{(n)},J_2^{(n+1)}))=\max_{J_1'}Q((J_1',J_2^{(n+1)})\mid(J_1^{(n)},J_2^{(n+1)})),$$

称之为 ECM 算法。一般地，ECM 算法比 EM 算法达到稳定的时间长.

另有一种 ECME 算法 (混合算法)，它就是交替地使用 ECM 算法和 EM 算法．

*2．在数据不完全时，用增补潜在数据，对参数的 Bayes 分布作估计 – Tanner–Wong 的潜变量法

２．１　基本想法－估计后验分布

在数据资料不全（缺失数据），或观测到的资料$\vec{Y}$并不是状态变量时，估计状态变量$\vec{X}$的分布密度$f(J,\vec{x})$中的未知参数J，与古典统计不同处是只能用观测到的资料$\vec{Y}$，这时可以通过后验分布密度$p(J\mid\vec{Y})$得到参数J的 Bayes 估计．然而，后验分布$p(J\mid\vec{Y})$一般并不知道，就需要用观测到的资料$\vec{Y}$对后验分布$p(J\mid\vec{Y})$进行估计，这就是本段的目的．

Tanner – Wong 的想法是：　在某些情形下，由观测数据$\vec{Y}$可以通过条件分布取样的机制，来构造某种"增补数据"（记为$\vec{Z}$，称为潜变量，Latent Variable)的样本值．于是$\vec{Z}$的这些样本值取自条件分布密度$p(\vec{z}\mid\vec{y},J)$．　$p(\vec{z}\mid\vec{y})$称为预测分布．而在$\vec{Y}=\vec{y}$已知的条件下，潜变量的各个样本值彼此是条件独立的．

潜变量$\vec{Z}$是观测不到的，如何选好它，最为重要．　最简单的情形是：　观测数据$\vec{Y}=Y_1,Y_2,\cdots,Y_N$为$N$个时刻的历史资料，其中$Y_i$是一个$m$维向量，　如果它的第１个分量是缺失的，而其它$m-1$维就是状态变量．这时候潜变量就可取为那个缺失的一维变量．

然而，一般情形远非如此简单．　潜变量的选取是关系到能否有效地计算的关键．　Tanner-Wong 选取$\vec{Z}$的原则是，同时满足以下两个条件（注意在上面所提到最简单情形，下面的条件是满足的）：

（1）易于模拟条件分布 $p(\vec{z} \mid \vec{y}, J)$ 随机数，

（2）容易计算后验分布密度 $p(J \mid \vec{y}, \vec{z})$.

下面的讨论都基于（1)，（2）均能够施行.

2．2　未知参数的后验分布的迭代估计

一旦选定潜变量以后，利用广义全概率公式，就得到

$$p(J \mid \vec{y}) = \int p(J \mid \vec{y}, \vec{z}) p(\vec{z} \mid \vec{y}) d\vec{z} \tag{15.6}$$

和

$$p(\vec{z} \mid \vec{y}) = \int p(\vec{z} \mid \vec{y}, J^*) p(J^* \mid \vec{y}) dJ^* . \tag{15.7}$$

这两个关系式是设计迭代算法的基点.

（迭代可能性的理论依据粗略地为： 把 (15. 7)代入(15. 6)得到

$$p(J \mid \vec{y}) = \int [\int p(J \mid \vec{y}, \vec{z}) p(\vec{z} \mid \vec{y}, J^*) d\vec{z}] p(J^* \mid \vec{y}) dJ^*$$

$$\overset{\text{记为}}{=} \int K(J, J^*) p(J^* \mid \dot{y}) dJ^* .$$

这说明 $p(J \mid \vec{y})$ 是积分核 $K(J, J^*)$ 的不变函数．积分核 $K(J, J^*)$ 依赖于观测值 $\vec{y}$ ，但是观测值 $\vec{y}$ 是固定的，所以没有将它计入记号. 只要对积分核 $K(J, J^*)$ 作适当的假定， 就可以使迭代算法收敛到积分核的不变函数).

由于积分核是不知道的，我们就不可能直接利用以下的迭代算法

$$g_{n+1}(J) = \int K(J, J^*) g_n(J^*) dJ^*$$

近似其不变函数． Tanner – Wong 提出用 Monte Carlo 迭代方法给出 $p(J \mid \vec{y})$ 的一个估计.Tanner – Wong 设计的迭代算法是，利用按预测分布的多个独立取样，得到潜变量的多个样本值， 通过它们由后验分布 $p(J \mid \vec{y}, \vec{z}^{(i)})$ 们更新未知参数 J 的后验密度 $p(J \mid \vec{y})$ 迭代估计. 注意(15. 6)说明 $p(J \mid \vec{y})$ 是 $p(J \mid \vec{y}, \vec{z})$ 关于预测分布密度 $p(\vec{z} \mid \vec{y})$ 的数学期望。 因此如果得到了潜变量的多个样本值 $\vec{z}^{(1)}, \cdots, \vec{z}^{(m)}$ ，就可以作 $p(J \mid \vec{y})$ 的如下的估计

$$\hat{p}(J \mid \vec{y}) = \frac{1}{m} \sum_{i=1}^{m} p(J \mid \vec{y}, \vec{z}^{(i)}) . \tag{15.8}$$

对此还需要计算潜变量加入条件的后验分布 $p(J \mid \vec{y}, \vec{z})$.

于是，估计未知参数 J 的后验密度 $p(J \mid \vec{y})$ 就归结为以下的迭代程序:

(1)　置初值 $p_0(J \mid \vec{y})$;

假定已经算出了第 n 次迭代的 $p_n(J\mid\vec{y})$，那么第 $n+1$ 次迭代的 $p_{n+1}(J\mid\vec{y})$ 用以下的两个步骤轮番地修改得到

(2)　I (Imputation) 步骤 (得到潜变量的条件样本)： 先按 $p_n(J\mid\vec{y})$ 对 J_n 作随机取样，再按 $p(\vec{z}\mid\vec{y},J_n)$ 独立地取样 m 个: $\vec{z}^{(1)},\cdots,\vec{z}^{(m)}$ (m 充分大);

(3)　P (Posterior) 步骤 (用增补后验分布更新后验分布)： 令

$$p_{n+1}(J\mid\vec{y})=\frac{1}{m}\sum_{i=1}^{m}p(J\mid\vec{y},\vec{z}^{(i)}).$$

3 几种智能算法

3．1　背景

随着高性能计算机的发展，出现了一系列算法，如神经网络，模拟退火，遗传算法，演化算法，隐 Markov 模型，自适应算法等．它们大多出自仿真人的思维，不仅具有通用，稳健(robust)，简单及便于并行处理等优点，而且也有望成为将数值计算与语义表达，形象思维等与高级智能行为相联系的桥梁．事实上，把人的智能与思维仅仅理解为主要是逻辑思维，是有很大的片面性的．与逻辑思维相反的以因果思维与统计判决为代表的非逻辑推断,应该说在日常推断中是更为本质的．例如，在逻辑推理中，从 ”A 蕴涵 B” 这个命题能作的推理是：一旦 A 出现，就立知 B 发生．这只是一种简单思维．而人类的智能推断是更为高级的思维．其判断过程却往往是一个由 ”结果” 探求其 ”原因” 的相反过程．例如，医生诊断疾病时，常常是：由 ”疾病 A 有症状 B” 这一命题，从病人有症状 B，反过来推断病人有多大的可能有疾病 A．这是一种非逻辑的推断，含有不确定性，还有犯错误的风险．然而这正体现了人类的高级智能活动．同样的推理还有：法医尸检死因，侦探寻找作案人，天气预报，考古分析，矿藏探测等等．与之相联系的计算方法都可归入智能算法．

一般地，智能算法常具有以下的特点：

(1)　大都引入了人为的随机因素，因而其计算过程实际上是作随机过程的模拟．然而,这种人工噪声的作用不仅不起干扰的作用，而是相反地起正面的作用．使用它完全是为了控制计算的进行方向．

(2)　大都具有自适应机制，可以在计算过程中不断调整．

(3)　大都是针对通用的目标函数而设计的，并不采用具体问题具体处理的启发式方法．

(4)　不少算法是对高维及复杂情形设计的．它们在低维或简单的情形有时效果很差，有时也显得很笨拙．

3．2 决定性的人工神经网络

神经网络是一个由彼此具有相互作用的多个神经元联结在一起的一个系统．作为神经网络的整个系统的协同工作产生一个总的效应．这个总的效应是一种功能，它可以是对应于一定的输入时有一定的相应输出，也可以是一种记忆功能．神经网络的重点是功能研究．在功能研究的基础上，构造具有相互作用的粒子系统，称为人工神经网络，以仿真地起到成为模拟某种功能的功能器．在数学上，人工神经网络也可以用来构造具有某种给定功能的分片线性处理器．人工神经网络的操作一般具有黑箱的特点．神经网络方法的精神在于：

由从例子(即样品)学习中，归纳出规律．

因此，它并不需要知道数据来自的对象的机理．这是它的长处．当然，在规律较为明显的情形，它也就会显出不足．总体看来，可以说神经网络方法是有别于经典统计的一种另类统计方法．

人工神经网络可分为决定性的和随机的两大类．它们又各自包括各种不同类型的神经网络．这种不同的神经网络的构造，是根据要求它完成某种制定的特殊功能而设计的．而随机因素的引入，是为了使计算更稳定，减少往错误的方向进行的可能性．

决定性的人工神经网络,如前传神经网络与随机过程并没有关系．但是它是人工神经网络的基础，它有广泛的应用． 再则，通过它不仅可以了解人工神经网络的原理与功能，并且也可以更好地理解随机的神经网络．所以本书也采纳了决定性的人工神经网络． 以下我们先介绍决定性的前传人工神经网络，而把更为一般的决定性的递归神经网络放在随机神经网络中一并处理．

3.2.1 决定性的前传人工神经网络

1．前传网络的概念

最常见的人工神经网络是多层前传网络．这种网络可分为多层，其中每层有一些单元组成．一般分为 3 种层次，依次为：输入层(由接受输入数据的神经元组成，它们不受其他神经元的相互作用)，隐层和输出层．作为中间层的隐层，又可分为若干个子层(由 Kolmogorov 的理论结果，它等价于只有一个隐层的某个系统．然而在实际计算中，往往以采用多层更为方便，灵活)．所谓前传网络，就是信息只能由前向后传，只有前一层的单元对其后一层的单元可能有相互作用．每一个隐层或输出层都由许多感受子(perceptron) 组成．每一个感受子就是一个受到网络相互作用的神经元．在前传网络中，这种相互作用只允许来自上一层的神经元．

最典型的是具有3层的全连接前传网络，其隐层元，例如，可以取6个至8个神经元．网络的效果相当于分片线性近似．它是一种进行自适应学习的非线性处理器. 其典型的模式是如下的受门限限制的线性作用：设输入层为 m 个输入神经元 $\vec{x}=(x_1,\cdots,x_m)$，隐层有 l 个神经单元(体现为分量) $\vec{h}=(h_1,\cdots,h_l)$，而输出层为 n 个单元 $\vec{y}=(y_1,\cdots,y_n)$． 用权重 $w_{ik}^{(1,2)}, w_{kl}^{(2,3)}$ 分别表示输入元 i 给隐元 k 的作用(强度)，和隐元 k 给输出元 j 的作用. 通常取门限函数 g 为在 0 点的单位阶跃函数

$$g(x)=\begin{cases}1 & (x\ge 0)\\ 0 & (x<0)\end{cases},\qquad (15.9)$$

或 S 形阈值函数, 例如,

$$g(x)=\frac{1}{1+e^{-ax}},\ \text{或}\ \ g(x)=\frac{2}{p}\arctan(ax),\quad (a>0).\qquad (15.10)$$

于是输入层到隐层间的转递为

$$h_k=g\Big(\sum_i w_{ik}^{(1,2)}x_i-q_k\Big),\qquad (15.11)$$

而隐层到输出层的转移为

$$y_j=g\Big(\sum_k w_{kj}^{(2,3)}h_k-q_j'\Big),\qquad (15.12)$$

其中 q_k, q_j' 是为了方便而预先设置的门限常数, 称为阈值.．在 g 为在 0 点的单位阶跃函数

时，g 取值 1 对应于神经元处于激发状态，取值 0 就表示处于抑制状态．在g为S形函数时，对应神经单元取连续的状态值，这种假定有利于在一些近似运算中求导数．此外即使在神经元只取抑制和激发两种状态时，也可以设置抑制和激发以外的一些缓冲状态．

前传神经网络是一种最简单的非线性处理器，即一种分片线性处理器，其输入是输出的分片线性函数．也就是由样品确定参数的分片线性近似器 (把它与用样条函数(Spline)近似作比较，前传神经网络除了有输出对输入并不可微的缺点外，操作与运行都远为简单).

2．前传人工神经网络的建模与应用

神经网络主要用在建模，模式识别，图像处理，基因表达等领域．多层前传网络的使用可以分为两个相位：训练学习相位与运转相位(操作相位)．

学习相位就是通过对样品的加工计算，得到权重常数 $w_{ik}^{(1,2)}, w_{kj}^{(2,3)}$ 和 门限$q_k, q_j{}'$等参数的估计值，使得在这组参数估计值下，该神经网络能相对最好地执行我们冀望它完成的任务．

神经网络的学习方法分为有监督的学习与无监督的学习．多层前传网络大多采用有监督的学习．其中最广为应用的是误差后传法．有监督的学习是指我们可以得到一组给定输出的输入样品．在学习时可利用已知的输出参照比较．学习的目的就是找出最佳未知参数，使得当将样品的输入部分输进这组参数下的网络时，能使输出最接近样品相应的输出．其要点就是根据已知的典型输入和输出，来估计连接系数 $w_{ik}^{(1,2)}, w_{kl}^{(2,3)}$，而其实质是灵活地运用最小二乘法.

误差后传法的主要步骤是：

(1)．设定参数的初始值；

(2)．逐个或按随机顺序将样品的输入，输进初始参数所对应的网络，得到输出，并将其与该样品应有的已知输出进行比较；

(3)．校正最后一层(设为第n层)的网络参数，得到使得输出最接近于应有输出的最后一层参数，并解出使得误差最小的第$n-1$层输出；

(4)．将第$n-1$层当作第n层，重复(3)，校正第 *n-1* 层的参数，并求出第$n-2$层输 出，使得第$n-1$层的输出误差最小；

(5)．重复(4)，向前面各层推进,直到求出所有各层的参数；

(6)．重复步骤(2)至(5)，直到看起来稳定．

这个算法是在未知状态(各个神经元的状态)的情况下估计参数，这也是缺失数据的一种参数估计．在估计参数的优化过程中，由于目标函数是非线性的，且变量数又十分大，计算往往会陷于局部极值，因此，也可考虑应用模拟退火等整体优化方法．

另一个相位是运转相位．它是在学习相位的基础上(就是在确定参数以后)，神经网络中各结点(即神经元)的状态按规定的方式运行(即进行运算)，以执行希望达到的任务，如预测，识别，分类等目的.

前传网络在计算机科学，人工智能等领域中有广泛的应用，它是一种比较好的另类回归模型．

3.2.2　一般的决定性的人工神经网络模型(神经元未必只取 1 或 0)

在一般的情形，神经网络对神经元x的总相互作用可以表示为

$$h(x) = g(\sum_y w(x,y)h(y) - q(x)) ,\qquad (15．13)$$

其中 $w(x,y)$ 是神经元 y 对神经元 x 的相互作用势，$h(y)$ 是神经元 y 所处的状态，而

$q(x)$ 则是神经元 x 的激发门限值．在神经元只取０或１（抑制或激发）两种状态时，乘积

$w(x,y)h(y)$ 是神经元 y 对神经元 x 的作用力，g 为门限函数，它是对自变量的总相互作用超过门限的部分起的作用，其中的和号取遍一切对于神经元 x 有相互作用的神经元 y 的全体．

３．３　随机的人工神经网络

3.3.1　一般概念

在随机的神经网络中，下一层神经元 x 被激发的概率取为

$$p(x)=\frac{e^{b\cdot h(x)}}{\sum_{x} e^{b\cdot h(x)}}，\ b>0\ . \qquad (15．14)$$

显见在 $h(x)$ 越大处，$p(x)$ 也越大，即该处的神经元 x 被激发的概率越大． 随机的神经网络与相互作用粒子的随机系统非常类似，它的运转是一个 Markov 链．

随机因素的加入，在下面的递归网络中能起控制计算进行方向的稳定作用．

3.3.2　递归(或反馈)(recurrent)网络与 Boltzman 机

递归(recurrent)网络也称为反馈网络，它实际上是以全体神经元所处的状态(称为一个组态)为变元的一个动力系统，因而也具有反馈的连接．在数学上反馈提供了迭代的功能，所以，同样作为离散动力系统，递归网络比前传网络具有更大的非线性功能，因而有更强大的应用潜力．递归网络也可以是决定性的，也可以是随机的．随机情形的网络，是按一个随机的规则发展的，因此，全体神经元所处的状态的时间发展，就是一个 Markov 链．

１．Hopfield 网络

Hopfield 把网络与吸引子联系起来，由此给出了一种联想记忆的模型，并给出了一种学习方法，通过学习训练确定网络连接参数，以能够使它达到所希望有的功能．Hopfield 网络是一种典型的递归网络．它源自人脑机制的研究．Hopfield 提出这种网络是希望用数学模型仿真人脑的功能．这种网络具有反馈的连接，相当于一个离散的动力系统．它不像前传网络那样简单地对输入与输出的关系感兴趣．决定性的 Hopfield 网络，一般地没有输入与输出，在给定初值后，就可以通过网络的运转不断地自动更新，最后落到网络系统的某些不变集上．在这意义下可以说并没有真正的输出，但是可以通过某些手段观测到它的自动更新的进程．

(1) 决定性的 Hopfield 网络

神经网络是一个由 N 个神经元彼此由相互作用而联结在一起的一个系统，整个系统在协同工作下产生一个总体效应．Hopfield 网络的总体效应，是提供记忆的数学模型，即联想记忆模型．设每个神经元可以取 -1 与 $+1$ 两个状态之一，其中 -1 代表该神经元处于抑制态，$+1$ 代表该神经元处于激发态。把 N 个神经元处的整体所处的状态的全体记为 X ，即

$$X=\{(x_1,\cdots,x_N):x_i=-1\text{或}+1,(i\le N)\}\ .$$

设神经网络对神经元 x 的总相互作用是(它是(15. 13) 的特殊情形)

$$y_i=\mathrm{sgn}(\sum_{k=1}^{N} w_{ik}x_k-q_i)\ , \qquad (15.\ 13)'$$

其中与神经元 i 相连接的神经元 k 有一个连接系数 w_{ik}，并且假定满足 $w_{ik}=w_{ki}$.

学习位相就是要估计这些参数，可以用 EM 算法.

运转相位有异步动力学与同步动力学两种不同的方式．Hopfield 的异步动力学发展是每次只容许依次地改变一个神经元，其具体步骤为：

(A) 先按(15.13)’ 把x_1更新为y_1. 并用y_1取代(15.13)’ 右方的x_1;

(B) 把x_2更新为用(15.13)’计算出的新的y_2，并用此y_2取代(15.13)’ 右方的x_2;

(C) 再把x_3更新为用(15.13)’计算出的新的y_3，并用此y_3取代(15.13)’ 右方的x_3;

(D) 如此依次地，一个一个地使网络中的所有N个神经元都更新一遍;

(E) 重复(A)至(D).

Hopfield 同步动力学发展与 Hopfield 异步动力学发展的不同之处是，不限于每次只改变一个神经元．同步动力学的时间发展，是通过(15.13)’把网络的状态从$\vec{x}=(x_1,\cdots,x_N)$转移为$\vec{y}=(y_1,\cdots,y_N)$．把它改写为通常的离散的动力系统就是

$$\vec{x}^{(n)}\overset{\Delta}{=}(x_1^{(n)},\cdots,x_N^{(n)})\to\vec{x}^{(n+1)}\overset{\Delta}{=}(x_1^{(n+1)},\cdots,x_N^{(n+1)}):$$

$$x_i^{(n+1)}=\mathrm{sgn}(\sum_{k=1}^{N}w_{ik}x_k^{(n)}-q_i)\,. \qquad (15.\ 13)''$$

同步动力学常常比异步动力学有更多的不变集.

递归神经网络更接近于生物神经网络．特别是 Hopfield 网络，它提供了一种联想记忆模式，即把神经网络动力学发展的一个不动点，或一个不变集（统称为神经网络的吸引子），解释为一个记忆．这种记忆模式具有与传统的地址记忆模式(如笔记本,硬盘，光盘等)完全不同的机制．它的记忆恢复不需要利用存入记忆的目录索引，而是按内容存取，因而它是迄今为止最接近于脑记忆功能的网络模式．当然为了增加新的记忆，就必须不断地更新整个神经网络，使得它能容纳新的吸引子.

（2）随机的 Hopfield 网络

- 随机的 Hopfield 异步神经网络

随机的 Hopfield 异步网络，是按照类似于随机场那样给出相互作用的．神经网络的时间发展是随机的，即按照一个 Markov 链x_n作随机转移，其中整个网络的状态(组态,记为$\vec{x}=(x_1,\cdots,x_N)$）之间的转移概率，例如可以取为

$$p_{\vec{x},\vec{y}}\overset{\Delta}{=}P(x_{n+1}=\vec{y}\,|\,x_n=\vec{x})=\begin{cases}\dfrac{1}{N} & (\vec{y}=(x_1,\cdots,x_{i-1},y_i,x_{i+1},\cdots,x_N),i\le N)\\ 0 & (\text{其它情形})\end{cases},$$

此处$y_i=\mathrm{sgn}(\sum_{k=1}^{N}w_{ik}x_k-q_i)$，$(i=1,\cdots,N)$.

注意，异步 Markov 发展的要义在于，转移矩阵是 Gibbs 型的，即转移时在网络的N个

神经元中只容许改变一个神经元．

- 随机的 Hopfield 同步神经网络

同步 Markov 发展的要义是，按转移矩阵转移时，在网络的N 个神经元中可容许同时改变所有的神经元．在 Hopfield 同步神经网络中的时间发展也是随机的，按照一个 Markov 链x_n作随机转移．整个网络的状态(组态)之间的转移矩阵中，还常常引进了一个倒温度参数 b ．转移矩阵则可以由下式给出(假定$\sum_{k=1}^{N}(w_{ik}x_k-J_i)\neq 0$)

$$p_{\vec{x},\vec{y}}(b)\stackrel{\Delta}{=}P(x_{n+1}(b)=\vec{y}\,|\,x_n(b)=\vec{x})=\prod_{i=1}^{N}\frac{1}{1+e^{-b(\sum_{k=1}^{N}(w_{ik}x_k-J_i))y_i}}．\quad (15.15)$$

注意，在 $b=\infty$ 时，$\vec{x}$就概率为 1 转移为$\vec{y}=(y_1,\cdots,y_N),y_i=\sum_{k=1}^{N}(w_{ik}x_k-J_i),(i\leq N)$，这就成为确定性的网络．这个 Markov 链与 Ising 模型的 Glauber 动力学非常相像．容易检查它有配称分布,且其配分函数为

$$Z_b(\vec{x})\stackrel{\Delta}{=}\sum_{\vec{y}}e^{\frac{b}{2}\sum_i\sum_{k=1}^{N}(w_{ik}x_k-J_i)y_i}．$$

所以，此 Markov 链有可逆不变分布

$$p_b(\vec{x})\stackrel{\Delta}{=}\frac{Z_b(\vec{x})}{\sum_{\vec{x}}Z_b(\vec{x})}．$$

决定性 Hopfield 同步网络的动力学发展的吸引子，就是(15.13)”的不变集，也就是方程

$$x_i=\mathrm{sgn}(\sum_{k=1}^{N}w_{ik}x_k-q_i)\qquad(15.\ 16)$$

的解$\vec{x}$．可以证明，对于随机的 Hopfield 网络的 Markov 链x_n发展而言，只要适当地控制倒温度 b ，就可以使 Markov 链x_n以大概率相对地留在方程(15.16)的某个指定的解附近．这就使随机 Hopfield 网络能够很好地解释联想记忆的机制）．

决定性 Hopfield 网络的联接系数的学习与更新

前面我们已经说过，Hopfield 网络的功能是用描述其网络发展的离散动力系统的吸引子给出的联想记忆模式．如何对 Hopfield 网络确定个联接系数$\{w_{ik}\}$使它能容纳给定的记忆(即吸引子)呢 ？这就需要对联接系数进行学习更新．Hopfield 给出了如下的学习律：如果要使$\vec{y}$成为吸引子，只要把原有的联接系数$\{w_{ik}\}$更新为新的联接系数$\{w_{ik}'\}$，其中

$$w_{ik}' = w_{ik} + c(2y_i - 1)(2y_k - 1).$$

但是,在此模型下，更新联接系数后的网络并不一定保留原网络的吸引子,除非假定各次加进的吸引子是彼此独立的． 所以用 Hopfield 网络来描述记忆远非完善，有很宽广的研究余地．

2．Boltzman 机

在 Hopfield 同步动力学网络中再加进输入神经元与输出神经元，就得到一种更为广泛的神经网络模型，称为 Boltzmann 机．Boltzman 机是一种神经网络，整个网络的时间发展也是一个 Markov 链，一般设置了输入．Boltzmann 机也可以看成是前传网络与 Hopfield 网络的结合与推广，它并不对输入层神经元加以“没有相互作用”的限制，也不限制相互作用只能由前向后．它又设置了输出神经元，同时它还是一个递归式的网络．所以它兼有两种网络的优点，可用以描述和模拟十分复杂的系统，具有很强的功能．

Boltzman 机的学习方法也是用 EM 算法．由于 Boltzman 机计算量非常大，学习起来很复杂，用来解决实际问题远不如前传多层网络普及，但应用潜力却高于后者．

Boltzmann 机输入神经元的状态由每个时刻的瞬间输入决定，而输出神经元是可观测的．设整个网络有 s 个输入单元，m 个隐单元，r 个输出单元.若以

$$I_k(n),(k \le s), h_l(n),(l \le m), O_j(n),(j \le r)$$

分别表示输入神经元，隐神经元与输出神经元在时刻 n 的取值．在

$$\vec{I}(n) \overset{\Delta}{=} (I_1(n), \cdots, I_s(n)), \vec{h}(n) \overset{\Delta}{=} (h_1(n), \cdots, h_m(n)), \vec{O}(n) \overset{\Delta}{=} (O_1(n), \cdots, O_r(n))$$

已知条件下，$\vec{h}(n+1), \vec{O}(n+1)$ 分别取值 $\vec{y} \overset{\Delta}{=} (y_1, \cdots, y_m), \vec{z} \overset{\Delta}{=} (z_1, \cdots, z_r)$ 的条件概率取为：

$$P(\vec{h}(n+1) = \vec{y}, \vec{O}(n+1) = \vec{z} \mid \vec{I}(n), \vec{h}(n), \vec{O}(n))$$

$$= \prod_{l=1}^{m} \frac{1}{1 + e^{-bU_l(\vec{I}(n), \vec{h}(n), \vec{O}(n)) y_l}} \prod_{j=1}^{r} \frac{1}{1 + e^{-bU_{m+j}(\vec{I}(n), \vec{h}(n), \vec{O}(n)) z_j}}, \qquad (15.17)$$

其中

$$U_i(\vec{I}, \vec{h}, \vec{O}) = \sum_{k=1}^{s} w_{ik}^{(I)} I_k + \sum_{l=1}^{m} w_{il}^{(h)} h_l + \sum_{j=1}^{r} w_{ij}^{(O)} O_j, \quad (i \le m + r), \qquad (15.18)$$

对于 $i \le m$ 及 $i > m$ 分别表示网络对于隐单元与输出单元的作用位势．

可以看出如果 $\vec{I}(n)$ 是一个 Markov 链，那么 $(\vec{I}(n), \vec{h}(n), \vec{O}(n))$ 也是一个 Markov 链．

Boltzman 机比同步网络应用更广，也更适合于复杂问题的模式识别．这种网络的缺点是： 由样品的实测值来估计相互作用系数(联接系数) $\{w_{ik}^{(I)}, w_{il}^{(h)}, w_{ij}^{(O)}\}$ 的学习训练过于复杂，其计算量非常大，然而估计的效果比较好，而且很稳健(Robust)．

特别地，如果只允许 $\vec{I}$ 对 $\vec{h}$ 及 $\vec{h}$ 对 $\vec{O}$ 有作用，这样的 Boltzman 机就是随机的前传网络．当β=+∞时就退化为前传网络．

[注] 神经网络方法的隐单元数，连接与门限都可以调节，这种分片非线性近似具有非常大的灵活性．如果选取得当，运用得好，效果会非常好．有些人认为这种方法过于黑箱化，会影响效果，其实效果不好往往由于选取不够灵活，并非方法的问题．再则，在能量函数中

用 S 形函数(sigmoid), 比用两值函数(signal)有明显的优点．例如在作分类器时, 那些明显地属于某类的样本, 既需要让它们起作用, 又不能让他们过分起作用, 因此需要用上下限来限制它们, 而中间的 S 部分正是用以区分的重要部分, 因为分类器的重要处在于边界处的区分．

当离散样本量不大时， 作者的经验是，可以在每个样本附近的一个小球内人工地加入一些均匀随机样本，然后再分组．

３．４　演化算法，遗传算法

遗传算法(Stochastic Genetic Algorithm, SGA)，或演化算法 (evolution algorithm)，是仿生算法，两者并无本质差别，只是不同的流派采用的不同命名．这是借鉴自然界的进化过程而设计的．人们注意到自然界在长期的自然选择过程中，往往能够达到惊人的优化水平，就自然地想到模仿这一过程，进行优化处理.各种应用实例表明这种算法有较好的稳健性．

遗传算法是一种启发式的 Monte Carlo 繁衍算法．它基于自然选择原理和自然遗传机制, 模拟自然界中的生命进化过程解决有复杂目标的非线性优化问题，包括参数优化，自动程序设计等问题．这些问题的特点是：解空间是已知的，并且可以比较不同解的优劣性．但是，由于这些问题的结构一般比较复杂,用一般推理的方法可能很难进行求解.与一般算法不同，这种算法并不从分析问题出发，而是直接从解出发．先随机地的生成多个解，然后仿照自然界生物进化的思想, 对这些解作遗传，杂交与变异运算，并根据优胜劣汰的思想保留更好的，摈弃较差的解, 得到子代．再不断的重复执行这种操作, 直至得到的解令人满意为止．这种算法的优点是有智能性, 包括具有自组织，自适应，自学习性．它无需事先研究问题的全部结构．再则, 它具有并行性, 极适合于大规模并行运算．而其缺点在于计算量大．然而，如果问题的解的适应性函数的性质不好(例如, 对于密钥号码的搜索问题, 它只有一个解，其适应性为 1，其它状态的适应性全为 0), 这时用演化算法就不能有效地求解．

遗传算法的基本形式是通过群体更新样本(再抽样)，作为自然选择，再佐以对样本作小概率的变异，以最终达到整体优化的目的．简单地说，就是自然选择加以允许有小概率的变异．而它的其它变化形式，只是它在具体应用中作不同的修改而形成的．

遗传算法是一个群体优化过程,为了得到目标函数的最小(大)值,我们不是从一个初始值出发，而是从一组初始值出发进行优化．这一组初始值好比一个生物群体，优化的过程就是这个群体繁衍，竞争和遗传，变异的过程．所以，其主体算法在数学上可描述如下：

假定有一个单变量目标函数U，我们并不知道它的解析表达式，但是对于任意给定的x，我们通过实验可以测量到$U(x)$的值(也就是知道函数的取值表)．这时，$U(x)$的优化问题(求最大值或求最小值的位置)可以设计为如下的程序：

取N个个体，每个个体取一个初始值，把此N个值写成一个列向量$\vec{x}_0=\begin{pmatrix} x^{(1)} \\ \vdots \\ x^{(N)} \end{pmatrix}$．我们引入此$N$个个体的时间发展机制，即它们以下面的随机迭代映射发展

$$\vec{x}_{n+1}=f(\vec{x}_n)+\vec{w}_n\,,\quad f=\begin{pmatrix} f_1 \\ \vdots \\ f_N \end{pmatrix},$$

其中的干扰随机向量 $\dot{w}_n=\begin{pmatrix}w_{n1}\\ \vdots\\ w_{nN}\end{pmatrix}$独立同分布，称为变异．从$\vec{x}_n$到$\vec{x}_{n+1}$的发展称为第 n 代到第$n+1$代的演化．从演化可以得到一个 Markov 链.而演化的目的是实现目标函数的优化，希望通过它们的时间发展，最后进入目标函数的优化值(最小值)附近．

最简单的演化是按某种优生原则，在此N 个个体中选出M 个进行优化（竞争），即选取M 个分量，对选中的分量i，例如（假定 $f(x)=\begin{pmatrix}f_1(x)\\ \vdots\\ f_N(x)\end{pmatrix}$），令

$$f_i(x)=-\frac{U(x+h)-U(x)}{h}$$

(在知道$U(x)$ 的表达式时，则取 $f(x)=-U'(x)$．类似地，在求最大值的问题中，相应地用$-f_i$ 代替 f_i)．而对没有选中的分量 j，则令

$$f_j(x)=x, w_{nj}=0\ .$$

这种迭代方案源自生物遗传的思想： $x\to f(x)$ 的转变，称为从父代到子代的遗传．用随机迭代映射的理论，在很宽的条件下，$\vec{x}_n$ 有一个极限分布$F^{(\infty)}(x)$，这个分布相对地更多地集中于目标函数的优化值附近．但是，这还不能完全达到求$U(x)$ 最小值的要求．求 $\min U$ 的位置有多种方法，其中之一是第 8 章中介绍的模拟退火方法．另外可以再加入父代与子代间的竞争机制，称为遗传竞争，即令

$$\vec{h}_{n+1}=\begin{pmatrix}h_{n+1}^{(1)}\\ \vdots\\ h_{n+1}^{(N)}\end{pmatrix}=f(\vec{x}_n)+\vec{w}_n\text{，}\ \vec{x}_n=\begin{pmatrix}x_n^{(1)}\\ \vdots\\ x_n^{(N)}\end{pmatrix},$$

其中

$$x_{n+1}^{(k)}=\begin{cases}h_{n+1}^{(k)} & (\text{若}U(h_{n+1}^{(k)})<U(x_n^{(k)}))\\ x_n^{(k)} & (\text{若}U(h_{n+1}^{(k)})\ge U(x_n^{(k)}))\end{cases}.$$

上面讨论的是单变量目标函数的优化．在目标函数是多变量函数时完全类似．在$M=1$且 f 满足一定限制条件下，用较多的概率论知识，就可以证明：在这种竞争机制下，当$n\to\infty$时，演化算法最终在概率收敛意义下，趋向目标函数取最小值点的某一个．

于是我们可以看到，遗传算法常常包含的机制与步骤为：

(1) 参数选择（设置初始值）：通常是对参数进行二进制编码工作．就是将空间中的一个点对应于一长串的单变量，对应于生物学中的每个”染色体”．这个二进制染色体的每一个二进制位称之为一个”基因”，基因的不同决定了染色体的不同．（需要注意的是，参数编码并不意味着仅仅从十进制换成二进制这么简单，不同的参数根据取值范围和精度的不同，

变换方式也不同)．初始群体模型(即初始值)应该是随机产生的，其个体在模型空间中分布越均匀越好，以使模型个体有足够多的遗传物质，以便能较好的找到整体极值．然而,群体数量过大会导致计算量的激增．

（2）竞争遗传(即自然选择)： 竞争选择是产生新群体（子代）的第一步． 应证优秀的个体的繁殖机会更大，这就是优生原则，选择初群体（父代，即初始值组)中的若干个个体来产生下一代．例如，可以把目标函数作为评估函数（这时的优化是取最大而不是取最小)，根据目标函数值的大小决定个体被选中的概率，并按这个概率选择初始群体中的个体，以体现优生原则．具体地，竞争机制可以通过用 Bootstrap 思想再抽样(Re-sampling)来实现． 例如说，对每个个体（ 即函数定义域 (是有限集) 中的x） 给予一个标志其优化程度的评估函数(其归一化就是权重，可以看成概率)，让具有更大的评估函数的个体以更大的可能性产生后代．例如，在求有限集上的函数 $f(x)(\geq 0)$ 的最大值的位置时，对 x 可取评估函数为 $f(x)$．并设置一个门限 e，把 $f(x)<e$ 的那些 x 先淘汰出局．而把留下的 x 们，按 $Cf(x)(C=(\sum_{f(x)\geq e} f(x))^{-1}$ 是归一化常数)作重要度抽样，独立地取样 N 次(N 充分大)，得到 N 个样本．假定其中样值 x_i(一共只有有限个)出现的次数为 $N_i,(N_1+\cdots+N_L=N)$．再把较小的 N_i 们淘汰出局，得到 ”遗传” 的子代(与一般的优化方法类似，自然选择一般只能保证达到局部优化．而加上杂交和变异，能得到整体性的改进)．

（3）杂交 (Crossover)：指从亲本模型重组产生新的子体模型的过程．在遗传算法中， 自然选择和杂交是最重要的部分，真正体现了遗传的特点．杂交就是遗传算法的一种繁殖过程．通过杂交可扩大当前的样本空间．选取亲本进行杂交的方式很多，一般可以设置一个交换概率 P_c，并选取一个亲本对． P_c越大说明要杂交的亲本对越多．还有一种方法是随机地从模型中选择亲本对进行杂交．杂交的实质是在模型空间进行更大范围的搜索，搜索的空间区域可能与原来的采样区域相当的远．所以，这种搜索是属于非邻近区域的(非局部的)搜索过程．它可以产生一个很有效的模型空间普查．一种常见的杂交方法是，对于两个非常高维数的离散数据(即被选中的个体)，决定性地或随机地取这两个数据的相同长度的一段分量，互相交换接成两个新个体，形成杂交的子代(对低维数据也可以便同地使用这种想法，例如，在一维情形可以把两个选中的个体 x, y 分别展开为二进小数列，决定性地或随机地取这两个序列中相同长度的一段互换，接成两个新个体)．一般地，针对不同的实际问题，我们可以根据对所需优化的问题的了解，选取其杂交方式．

(4) 变异：变异是偶然地对后代中的一个随机选择的基因作随机摄动．这是一个低概率的随机变化过程(即设定的变异概率 P_m 很小)．如果不存在变异，子代模型就永远得不到初始群体中没有的染色体基因，因而很难产生强有力的进化．变异的方法就是将某一个参数位的值，从 1 变为 0 或从 0 变为 1．即在优化过程中以小概率不按局部优化的方向进行．从局部看，这种演化似乎反而“ 劣化 ”了目标函数．但是，正是这种演化才为逃逸出局部极值的陷阱提供了可能性．由于我们可以控制“ 劣化 ”的变异只以很小的概率发生，在计算进行了很多代后，总趋势会向优化发展的．这种做法的基本思想与模拟退火算法是类似的．由于杂交和变异的存在，使遗传算法在模型空间作搜索中有很大的潜力．

(5) 更新：就是根据自然界中的”适者生存”的原则，取子代模型中的一个与其亲本比

较，摈弃适值（评估函数）小的亲本或子本．这是父代和子代之间的竞争．

(6) 算法停止规则：在多次进行（２）至（５）(注意，这些步骤并不要求按次序进行）之后，模型群体经过多次的竞争选择，杂交，变异和更新，群体的平均评估值（目标值）逐渐变大，直至聚集在模型空间的一个小范围之内．这时，可以用多种判别法来判断算法是否已经达到稳定．例如，可以用群体的平均目标函数值不变，作为算法可以停止的标准．也可以用群体的最大目标函数值与平均目标函数值是否接近，作为算法可以停止的标准等等．

总之，遗传算法是一组变量一起演化，所以是群体的优化，它特别适合于并行计算．另外，群体的规模也可以不是恒定的．

4 聚类，Kohonen 自组织学习，自适应算法

４．１ k－平均 (k -mean) 聚类

概括地说，聚类问题就是根据一些样品,寻找较少数目的代表点(在模式识别中就是建立一些"标准" 模板的过程)．这些样品称为学习样品或训练样品．一般地，样品常具有很高的维数，而高维空间是非常分散的．样品的总数一般远远小于高维区域的离散采样能近似地给出直方图所需的数目．于是这种统计问题与经典的统计有很大的不同，这时，一个给定的子区域能包含样本点的可能性一般非常小．但是，如果样品数据集能很好地表达它的总体，一般地数据就会集中在一个体积相对地小的一个区域．于是聚类问题的重要点就是去发现这个分布的总体位于那一个相对小的区域．

k－平均聚类就是预先设定为 k 个类，从样品寻找 k 个类的代表点．其想法是：先随机地找 k 个点作为代表点的初值，按与这些初值接近的程度，把所有的样品点分成 k 类，然后求每个类的平均值，分别作为初值的修改值．再重复以上的步骤，直至稳定为止．

k－平均聚类的优点是想法简单，设计方便．其缺点是缺少稳健性，对初始值的取法与初始类的数目的设置都过于敏感． 在这一方面，自适应聚类的效果要明显得多．

４．２ 自适应聚类的基本思路

自适应聚类就是预先并不设定聚类的个数．而是根据学习样品来得到"标准的"代表点(或模板)．这就是学习或训练问题，其想法是，找到数目尽量小而能覆盖所有数据的一组小球 (这与数据压缩中的矢量量化有点类似)，使每个球只包含一个模式的数据，这个球的中心就可以作为一个代表点(或模板)． 随之，统计分析就可以只限在每个小球上进行．而对于给定的样本的识别问题，就是要在代表点中找出这个被判别样本的合适代表．所以，模式识别就归结为决定被判定样本属于那一个球的问题．

自组织神经网络是对于具有固定规模的学习训练问题的改进．这是一种自动确定代表点的规模大小的自适应聚类，其基本思路是：在聚类过程中，采取竞争学习的机制．这种想法对于模板的样品集为非常分散的超高维数据的情形比较适用．如果一个样品在现有的模板中实在找不到合适的代表，那么，就用此样品作为一个新的模板．

４．３ 固定规模的 Kohonen 网络

Kohonen 的自组织拓扑映射(Kohonen’s Topological Mapping，KTM) 是一种功能非常强的，用作聚类的网络．但是其计算量非常大，网络的规模的确定也比较困难．然而，在一定条件下它可以改变为自动确定网络规模的算法．用它作计算机科学中的“矢量量化”(相当于聚类，即选一些不依赖坐标的代表点，每个代表点作为一个“向量”)是很好的工具．KTM 的特长在于它要求有所谓的“拓扑性”，其含意是，用该网络所得到的代表点组能保持标号的序之间的拓扑关系不变，也就是：具有较近序标的两个代表点也较近．

KTM 的计算步骤是：设样品集为 $\{\vec{s}_1,\cdots,\vec{s}_m\}$

（1）．根据选取好的规模（即代表点总数，记为 N ），随机地或决定性地选取一组初值：$\{\vec{x}_0(1),\cdots,\vec{x}_0(N)\}$ （一般每个 $\vec{x}_0(i)$ 都是一个与每个样品相同维数的向量）；

（2）．随机地(即独立同分布地，也可以按某一决定性的规则)从样品集 $\{\vec{s}_1,\cdots,\vec{s}_m\}$ 中读出一个样本 $\vec{y}_n$；

（3）．竞争修改待选的代表点：设 j 满足

$$d(\vec{x}_n(j),\vec{y}_n)\le d(\vec{x}_n(i),\vec{y}_n)，(\forall i=1,2,\cdots,N)，$$

则令

$$\vec{x}_{n+1}(i)=\vec{x}_n(i)+a_n\cdot r(d(i,j))(\vec{y}_n-\vec{x}_n(i))，$$

其中 $r(r)$ 是一个严格单调递减的正函数，满足 $r(0)=1$，a_n 是一个使算法收敛的因子，称为冷却因子，而 $d(i,j)$ 则表示 i,j 间的距离；（即其序标与和读入的样本最近的代表点的序标相近的代表点，在修改时更多地向读入的样本靠拢）．

（4）．重复(2)和(3)，直至稳定．将稳定的结果记为：$\{\vec{x}(1),\cdots,\vec{x}(N)\}$．那么，$\vec{x}(1),\cdots,\vec{x}(N)$ 就是最后选取到的 N 个代表点．

Kohonen 猜测，在条件

$$\sum_n a_n=+\infty，\sum_n a_n^2<+\infty$$

满足时，算法是收敛的． $r(r)$ 严格正能够使最后得到的代表点近似地有拓扑性，即除了较少数的例外代表点以外，对于多数 $i,j\le N$，如果 i 与 j 较近则代表点 $\vec{x}(i)$ 与代表点 $\vec{x}(j)$ 也较近．

然而，除了一维情形以外，对高维的 Kohonen 网络所达到的功能，至今还只有计算机结果的图示，尚缺乏清晰的论述，还谈不上算法收敛性的严格证明．

如果不要求代表点具有拓扑性，那么就可以不必要求 $r(r)$ 严格正，而只需 $r(r)\ge 0$，即它可以取 0．这表示在更新待选代表点的时间进程中，有些待选代表点可以在某些时刻不修改．其中最简单的是取 $r(r)=\begin{cases}1, & (r=0)\\ 0 & (r\ne 0)\end{cases}$，这称为 ”胜者统吃 ” 的原则，它表示只对与读入样品 $\vec{y}_n$ 最相近的待选代表点做修改，这种修改表现为往读入的样品 $\vec{y}_n$ 处拉近一点．此时竞争修改代表点的算法为：设 j 满足

$$d(\vec{x}_n(j),\vec{y}_n)\le d(\vec{x}_n(i),\vec{y}_n)，(\forall i=1,2,\cdots,N)，$$

则令

$$\vec{x}_{n+1}(i)=\vec{x}_n(i),(i\neq j)\text{，}\vec{x}_{n+1}(j)=\vec{x}_n(j)+a_n(\vec{y}_n-\vec{x}_n(j))\text{。}$$

若每次只修改一个代表点，则当代表点数很大时,这种算法要比标准的 KTM 算法要快得多．例如，当代表点数为 300，则按”胜者统吃” 的原则计算时， 实践表明每次只修改一个代表点的算法比标准的 KTM 算法要快 300 倍．此外,为了具有拓扑性,需要调整代表点的排序,所耗费的时间远超过选取待选代表点的时间． 文献中证明了”胜者统吃”算法的组态(全体待选代表点组成的向量)按分布收敛到一个分布(当然一般不具有拓扑性)．再则，从应用的角度看，给出区间估计往往比收敛性更为重要．

［注 1］ 按”胜者统吃”的原则来聚类，可以化为目标函数的最小值问题．设样品集为$\{\vec{s}_1,\cdots,\vec{s}_m\}$．要得到$N$（$m>>N$）个代表点，则代表点$\vec{x}_1,\cdots,\vec{x}_N$就是目标函数

$$U(\vec{x}_1,\cdots,\vec{x}_N)=\sum_{k=1}^{m}\ \min_{i\leq N} d(\vec{x}_i,\vec{s}_k)^2$$

取最小值之处．这个目标函数的含义为：对每一个样品$\vec{s}_k$，希望在N个待选代表点$\vec{x}_1,\cdots,\vec{x}_N$中选取一个与样品$\vec{s}_k$最近的，作为样品$\vec{s}_k$的代表．

［注 2］ 对于代表点的规模不甚清楚,而有可能出现冗余的待选代表点的情形,Jinwen Zhang 及 Lei Xu 从另一种不同的角度考虑： 对于读入的样品$\vec{y}_n$，让与它最近的待选代表点往$\vec{y}_n$拉一点，而让与它第二近的待选代表点往离$\vec{y}_n$远处推一点，而其它待选代表点则保持不变．这种想法是希望最后把冗余的待选代表点推向不起作用的遥远之处．此时，竞争修改代表点的算法可以为：如果j,k满足

$$d(\vec{x}_n(j),\vec{y}_n)\leq d(\vec{x}_n(i),\vec{y}_n)\text{，}(\forall i=1,2,\cdots,N)\text{，}$$

$$d(\vec{x}_n(j),\vec{y}_n)<d(\vec{x}_n(k),\vec{y}_n)\leq d(\vec{x}_n(i),\vec{y}_n)\neq d(\vec{x}_n(j),\vec{y}_n)\text{，}$$

那么，例如，可以令

$$\vec{x}_{n+1}(i)=\vec{x}_n(i),(i\neq j,i\neq k)\text{，}$$

$$\vec{x}_{n+1}(j)=\vec{x}_n(j)+a_n(\vec{y}_n-\vec{x}_n(j))\text{，}$$

$$\vec{x}_{n+1}(k)=\vec{x}_n(k)-b_n d(\vec{y}_n,\vec{x}_n(k))^{-1}(\vec{y}_n-\vec{x}_n(k))\text{。}$$

这种想法还相当于使用了类似于如下的目标函数：

$$U(\vec{x}_1,\cdots,\vec{x}_N)=\sum_{k=1}^{m}\ [\min_{i\leq N} d(\vec{x}_i,\vec{s}_k)^2+\min_{j\leq N,\,j\neq i_0,\,d(\vec{x}_{i_0},\vec{s}_k)=\min_i d(\vec{x}_i,\vec{s}_k)}\frac{1}{d(\vec{x}_j,\vec{s}_k)}]\text{。}$$

实证显示，用这种想法常比”胜者统吃”能达到更为好的效果．这里目标函数的含义为：对每一个样品$\vec{s}_k$的代表点的选取，不只考虑与之最近的代表点的作用，还要顾及与之第二近

的代表点的作用．及希望兼顾地修改与样本最近的代表点使之更近，修改与它第二近的代表点使之更远．

4．4．网络规模的竞争学习

Kohonen 映射的缺点在于:

(1) 可能使计算陷入某种局部极小．为此可以用加入随机干扰以逃逸出这类陷阱.

(2) 在 Kohonen 自组织算法中，必须事先确定代表点（小球）的数目．但是通常给定一组数量较大或维数很高的数据，往往不能知道取多少代表点合适．这样，就自然地提出了怎样由数据自动地确定代表点的个数的问题，这就是规模问题．改进的方法是使代表点的数目依赖于网络的结构．这就是 4.1 段中的自适应竞争学习的思想，其目标是：对于给定的合适的精度$e>0$，寻找数目小而能覆盖所有数据的一组半径为e的小球，把一个小球的中心作为一个代表点(在模式识别中代表同一类模式)．在样品的代表球内就可以运用传统的统计，决定数据是否能降维，数据在球内的分布等等． 在模式识别中，不同的球的联系还有助于知道模式的非线性结构的复杂程度．而代表点的规模的大小，就相当于提供了用分片线性函数近似所需要的分片数与需要使用的神经网络的规模的信息．

对于自适应竞争学习，Grossburg 等提出了适应共振理论（Adaptive Resonance Theory，简称 ART），其含义为，当一个新的样本输入到学习的算法程序中时，只有它和某个待选代表点充分接近，它才能对这个待选代表点进行修正．这样就造成一种自适应的谐振状态．ART 算法大致是这样：一开始选取代表点数很少或为 0；由接受的输入样本，按照预先设计好的相似度衡量方法与标志 ”充分接近” 的门限，按照上述竞争原则，决定是否应按这个输入修正某个待选代表点，以及修正哪一个待选代表点．如果没有一个待选代表点和这一新的输入充分接近，则把此输入增加为一个新的代表点．

这个算法有相当好的稳健性.目前文献中给出了两种相似度控制方法： e-邻域控制法与中心控制法．这是多种模式的模式识别样品选取代表点 (并不要求一个模式只选取一个代表点)的自适应方法．用这种算法对手写体数字模式识别的样品选取代表点(矢量量化)，取得了很好的效果．由于这种系列参数适应算法不要求事先给定网络规模和控制参数，从而，它可以用来作为探索性统计分析的工具．

总之，自组织学习网络的中心思想，是利用演化历史中获得的信息指导搜索或计算，其中还可以人为地引入干扰，以免算法被套入局部陷阱．它一般表达为一个自适应的迭代算法，其形式为

$$x_{n+1}=f(x_n,y_n,w_n)\text{，}$$

其中x_n代表系统在时刻n时的状态，y_n代表样品，w_n代表为使搜索过程收敛而人工加入的干扰．

5 适应最小二乘法 -- 一种适应的变步长的随机逼近

设线性回归问题的随机输入为$\vec{x}_n$，观测为y_n，干扰为w_n，待估参数为$\vec{J}$的模型为

$$y_n=\overrightarrow{x_n}^T\vec{J}+w_n\text{．}$$

对此可以设计如下的递推估计：

$$\hat{J}_{n+1}=\hat{J}_n+\overline{e}_n\dot{x}_n(y_n-\dot{x}_n^T\hat{J}_n)\text{，}$$

$$e_{n+1} = e_n + g_n \vec{x}_n (y_n - \vec{x}_n^T \hat{J}_n) ,$$

$$\bar{e}_n = \bar{e}_{n-1} + \frac{1}{n}(e_n - \bar{e}_{n-1}) = \frac{1}{n}\sum_{k=1}^{n} e_k ,$$

其中

$$\sum_k e_k < \infty, \sum_k g_n = \infty, g_n = O(\frac{1}{n}), \bar{e}_n \to 0 .$$

那么，可以证明 $\hat{J}_n$ 依概率收敛与某个 $\hat{J}$，它就作为 $\vec{J}$ 的估计. 与经典的最小二乘法相比，这个方法优点是： 它充分地利用了新的观测，递推地修改估计，并且最后达到概率收敛.

龚光鲁，钱敏平著　应用随机过程教程 – 与在算法和智能计算中的应用

清华大学出版社，2003

第16章　离散状态的Markov控制与决策过程简介

(Controlled Markov Process, Markov Decision Process, MDP)

1 例

1．1 随机决策模型的简单例子

定义１６．１　随机决策模型的对象是可以控制的随机系统，人们可以选取控制决策，以改变发展过程的路径．在任意固定时刻，系统随机地处在 $S=\{1,2,\cdots,N\}$ 中的某个状态，而在策略取定为 a 的情况下系统的发展是按照一个随机矩阵 P(a) 作为转移概率阵而变化．这就称为一个 Markov 决策过程．

从下面的简单例子，可以得到一些直观的认识。

例１６．２　设某个经营系统总处在"1","2","3"三种状态之一．假定在每个整值时刻可选择两种不同的动作之一：$a_{(1)}$ 或 $a_{(2)}$，而在采取动作 $a_{(1)}$ 或 $a_{(2)}$ 时，状态间的转移矩阵分别为

$$P(a_{(1)})=\begin{pmatrix}\frac{1}{2} & \frac{1}{2} & 0\\ 0 & \frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & 0 & \frac{1}{2}\end{pmatrix},\quad P(a_{(2)})=\begin{pmatrix}\frac{1}{2} & 0 & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2} & 0\\ 0 & \frac{1}{2} & \frac{1}{2}\end{pmatrix}.$$

假定开始时（即时间 $n=0$ 时）该系统以相等的可能性处在这三个状态之一，即初始分布为

$\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$．又设处在状态 i 时，采取动作 $a_{(1)}$ 能得到报酬为 $g(i,a_{(1)})=2i$，而处在状态 i 时，采取动作 $a_{(2)}$ 能得到报酬为 $g(i,a_{(2)})=i^2+\frac{1}{2}$．我们要在各个时刻，根据历史状况,有目的地选取动作 $a_{(1)}$ 或 $a_{(2)}$，使在时间区段 $0\le n\le m$ 内得到的平均累积报酬最大．这里，动作是历史状况的函数．从时刻 n 的历史状况到采取的动作的对应（即函数），称为时刻 n 采取的策略．各个时刻采取的策略合起来，称为一个策略．我们要选取一个策略，使在时间区段 $0\le n\le m$ 内得到的平均累积报酬最大．

把在时刻 n 采取的动作记为 a_n，那么它只能 $a_{(1)}$ 或 $a_{(2)}$ 之一．于是转移矩阵 $P(a_n)=(p_{ij}(a_n))_{i,j\le N}$ 有确切的含义．这样，由初始分布 $m_0=(m_1,m_2,m_3)=(\frac{1}{3},\frac{1}{3},\frac{1}{3})$ 及转移矩阵列{ $P(a_n)$ }决定了一个 3 个状态的非时齐的 Markov 链 $\{x_n:n\ge 0\}$．x_n 代表系

统在时刻 n 所处的(随机的)状态．于是系统在时刻 m 前所得的平均累积报酬为 $E(\sum_{n=0}^{m} g(x_n, a_n))$．它就是需要优化的目标函数．

动作 a_n 的选取，直接影响了在时刻 n 以后此 Markov 链的样本的走向 $\{x_{n+1}, \cdots, x_m\}$．一般地，动作 a_n 依赖于系统的发展历史，即依赖于 $\{x_0, a_0, x_1, a_1, \cdots, x_{n-1}, a_{n-1}, x_n\}$．这里我们简单地限制在时刻 n 所采取的动作只依赖于当时所处的状态，也就是假定 a_n 只是 x_n 的函数，即 $a_n = f_n(x_n)$（其中 $f_n: \{1,2,3\} \to \{a_{(1)}, a_{(2)}\}$，是根据所处的状态选取的动作，即在时刻 n 所采取的策略)．我们先以 $m = 1$ 为例，看如何求得最高的平均累积报酬．也就是让 $a_0 = f_0(x_0)$，$a_1 = f_1(x_1)$，要选取函数(映射) f_0, f_1，使

$$V(f_0, f_1) \overset{\Delta}{=} Eg(x_0, f_0(x_0)) + Eg(x_1, f_1(x_1))$$

取到最大值．注意

$$Eg(x_0, f_0(x_0)) = \sum_{i=1}^{3} g(i, f_0(i)) m_i \text{．}$$

而采取了动作 $a_0 (= f_0(x_0))$ 后，非时齐的 Markov 链 x_n 从时刻 0 到时刻 1 的转移矩阵应该是 P($f_0(x_0)$)，于是

$$P(x_1 = j) = \sum_{i=1}^{3} m_i p_{ij}(f_0(i)) \text{．}$$

从而

$$Eg(x_1, f_1(x_1)) = \sum_{j=1}^{3} g(j, f_1(j)) P(x_1 = j)$$

$$= \sum_{j=1}^{3} \sum_{i=1}^{3} m_i g(j, f_1(j)) p_{ij}(f_0(i)) \text{．}$$

也就是

$$V(f_0, f_1) = \sum_{i=1}^{3} g(i, f_0(i)) m_i + \sum_{j=1}^{3} \sum_{i=1}^{3} m_i g(j, f_1(j)) p_{ij}(f_0(i)) \text{．}$$

由此式看出，若要选取策略 (f_0, f_1) 使 $V(f_0, f_1)$ 的值为最大，需先选取 f_1 使 $g(j, f_1(j))$ 最大．对此我们观察到

$$(j = 1) \qquad g(1, a_{(1)}) = 2 > \frac{3}{2} = g(1, a_{(2)}) \text{，}$$

$$(j=2) \qquad g(2,a_{(1)}) = 4 < \frac{9}{2} = g(2,a_{(2)})\text{，}$$

$$(j=3) \qquad g(3,a_{(1)}) = 6 > \frac{19}{2} = g(3,a_{(2)})\text{．}$$

可见，要使 $g(j, f_1(j))$ **最大，** f_1 **应取如下定义的** f_1^* **：**

$$f_1^* : (1,2,3) \to (a_{(1)}, a_{(2)}, a_{(1)})\text{．}$$

将在确定了 $f_1 = f_1^*$ **后的最大报酬记为** g^* **，那么**

$$g^*(j) = g(j, f_1^*(j)) = \begin{cases} 2 & (j=1) \\ \dfrac{9}{2} & (j=2) \\ 6 & (j=3) \end{cases}$$

于是

$$V(f_0, f_1^*) = \sum_{i=1}^{3} m_i [g(i, f_0(i)) + \sum_{j=1}^{3} g^*(j) p_{ij}(f_0(i))]\text{．}$$

剩下的就是选取 f_0 **，使** $V(f_0, f_1^*)$ **最大．为此只需使方括号中的各个量都取到最大．与前面不同之处，只是用方括号中量代替了前面的** $g(j, f_1(j))$ **而已．下面我们列出它的值，由** $p_{ij}(a_{(1)})$ **与** $p_{ij}(a_{(2)})$ **的定义得**

	$g(i,a_{(1)}) + \sum_{j=1}^{3} g^*(j) p_{ij}(a_{(1)})$	$g(i,a_{(2)}) + \sum_{j=1}^{3} g^*(j) p_{ij}(a_{(2)})$
$i=1$	$2 + \frac{1}{2}(2 + \frac{9}{2})$	$\frac{3}{2} + \frac{1}{2}(2+6)$
$i=2$	$4 + \frac{1}{2}(\frac{9}{2} + \frac{19}{2})$	$\frac{9}{2} + \frac{1}{2}(\frac{9}{2} + 2)$
$i=3$	$6 + \frac{1}{2}(2 + \frac{19}{2})$	$\frac{19}{2} + \frac{1}{2}(6 + \frac{9}{2})$

．

比较其各行的大小，可知 f_0 **应取策略** f_0^* **：**

$$f_0^* : (1,2,3) \to (a_{(2)}, a_{(1)}, a_{)2)})\text{，}$$

其对应的值(最大值)分别为： $\frac{11}{2}$ **，** 11 **，** $\frac{59}{4}$ **．所 以**

$$V(f_0^*, f_1^*) = \frac{1}{3}(\frac{11}{2} + 11 + \frac{59}{4}) = \frac{125}{12}\text{．}$$

于是在我们所限制的策略类之中，最佳决策为由如上定义的 (f_0^*, f_1^*) **所确定的**

$$a_0 = f_0^*(x_0)\text{，}a_1 = f_1^*(x_1)\text{，}$$

而在$n \le 1$时段内的最高平均累积报酬为 $V(f_0^*, f_1^*) = \frac{125}{12}$.

1．2　简单模型的启示

由例１６．２可以看出，如果限制在形如$a_n = f_n(x_n)$的策略类中，去找最佳的策略（即″从状态到动作的对应" f_n，那么，只要先选定时刻最后的m时刻所对应最佳的f_m^*，然后向后归纳地选最佳的 $f_{m-1}^*, \cdots, f_0^*$． 由此可以抽象出第2节中较为一般的数学模型．

2　动作只依赖当前所处状态的简单决策模型

2．1　简单模型的一般描述

定义１６．３　（决策动作不依赖系统的状态的情形）

假定在参数a（$a \in$某个有限集A，称为行动集）固定时，$P(a) = (p_{ij}(a))_{i,j \le N}$ 是一个以$S = \{1,2,\cdots, N\}$为状态空间的转移矩阵．设在时刻0,1,...各选一个行动，记为 $a_0, a_1, \cdots (a_i \in A)$，那么由初分布$m_0 = (m_1, \cdots, m_N)$及转移矩阵序列{$P(a_n)$:n≥0}可以决定一个非时齐的Markov链$x_n$，满足：

$$P(x_0 = i) = m_i, \quad P(x_{n+1} = j \mid x_n = i) = p_{ij}(a_n).$$

假定时刻n系统处在状态i时，采取行动a_n得到的报酬由报酬函数$g_n(i,a)$表示,那么在时刻m得到的累计报酬为 $\sum_{n=0}^{m} g_n(x_n, a_n)$，其中$g_n(i,a)$是在时刻$n$采取行动$a$且处在状态$i$时的报酬函数，那么,平均累计报酬为$E[\sum_{n=0}^{m} g_n(x_n, a_n)]$.

定义１６．４　（决策动作仅依赖系统当前的状态的情形时的期望总报酬）

这也是一种简单情形，例１６．３　是它的特例． 这时容许a_n的取值依赖于链所处的状态i的情形，即 $a_n = f_n(i)$的情形，其中f_n是状态集S到动作集的A的一个映射，其含义为：若Markov链在时刻n处于状态i，则采取决策$a_n = f_n(i)$．令

$$P_n = (p_{ij}(f_n(i)))_{i,j \le N}, \qquad (16.1)$$

则它仍是一个随机矩阵．由初始分布m_0及{P_n,n≥0}决定了一个非时齐Markov链x_n．类似地由报酬函数$g(i,a)$可以得到时刻m的平均累计报酬．此Markov链x_n在各个时刻的转移

矩阵是不同的，它们依赖初始分布μ_0及各个时刻的策略映射序列$\{f_n, n \ge 0\}$．我们记

$$\mathbf{f} \stackrel{D}{=} \{f_n, 0 \le n \le m\}, \qquad (16.2)$$

并称它为一个策略．于是，使用它得到的平均累计报酬为$E_{\mu_0}\left(\sum_{n=0}^{m} g_n(x_n, f(x_n)\right)$. 注意，对于非时齐的 Markov 链$x_n$的轨道（$x_0$，$x_1$,$\cdots$，$x_m\}$而言，我们采取的行动列为$\{f_0(x_0),\cdots, f_m(x_m))$．由于我们的行动列只依赖于 Markov 链当前所处的状态，这样的特殊策略 $\mathbf{f} \stackrel{\Delta}{=} \{f_n = f_n(*), 0 \le n \le m\}$也称为 Markov 策略，这时动作$a_n = f_n(x_n)$是随机的．在 Markov 链的初分布为$\mu_0$时，我们将 f 在时刻$m$取得的平均累计报酬记为$J(\mu_0, \mathbf{f})$:

$$J(\mu_0, \mathbf{f}) = E_{\mu_0}\left(\sum_{n=0}^{m} g_n(x_n, f(x_n))\right). \qquad (16.3)$$

在系统的初始状态为i时，平均累计报酬为 $J(i, \mathbf{f}) = E_i\left(\sum_{n=0}^{m} g_n(x_n, f(x_n))\right)$. （有时在总报酬中，除了累计报酬外，还要加上一个终止报酬$h(x_m)$，此时

$$J(\mu_0, \mathbf{f}) = E_{\mu_0}\left(\sum_{n=0}^{m} g_n(x_n, f(x_n) + h(x_n)\right).$$

而其数学处理是完全一样的)．

［注 1］ 以上考虑的是纯策略，更为灵活的是使用混合策略，也就是随机策略，它以给定的概率分配取动作集A中的不同动作，抽象地可以看成一个取值于A的概率向量(概率分布)m．这时的动作集A就用取值于A的全体概率向量组成的集合(记成P)所代替．注意，我们可以认为$A \subset \mathrm{P}$，因为纯策略是一个特殊的随机策略．在随机策略类P中考虑累计报酬，其每一步计算都应作相应的改变．在使用随机策略时的最佳报酬函数相应地为$V_{0,m}(\cdot)$，其中

$$V_{k,m}(i) \stackrel{\Delta}{=} \sup_p \ E_i\left(\sum_{n=k}^{m} g_n(x_n, p_n)\right), \qquad (16.4)$$

而$p = (p_0, p_1, \cdots), p_n$表示时刻$n$使用的随机策略．可以证明，当$g_n(i,a) = g(i,a)$(不依赖$n$)时，$V_{k,n}$关于$k$满足一个 Bellman 型向后递推公式

$$V_{k,m}(i) = \sup_{a \in A}[g(i,a) + \sum_{j=1}^{N} p_{ij}(a) V_{k+1,m}(j)],$$
$$V_{m,m}(i) = 0$$

由此可以反向地逐步解出在随机策略下的最佳报酬函数. 显见，它不小于在纯策略下的最佳报酬函数.

在其它的情形, 例如下面将介绍的折扣模型的情形, 实际上只须沿着这个思路方向, 做相应的改变就也就可以得到最佳报酬函数. 在历史发展给定条件下，随机策略的条件分布如果只与系统的当前状态有关，则这样的随机策略称为 Markov 随机策略． 如果只与系统的初始状态与当前状态有关，则称为半 Markov 随机策略．

［注 2］ 在实际情形中，更多出现的是复杂的情形，即对不同的状态可采取的动作集也不同，就是说，如果系统处在状态i，则能选取的动作集合为A_i． 于是这时相应地要求$f_n(x_n) \in A_{x_n}$． 为了使读者领会实质，在本书中我们只讨论A_i们相同的简单情形，其实其方法与理论并无二致．

［注 3］ 以上考虑的是 Markov 策略，即当前时刻所采取的动作只依赖于当前所处的状态的简单情形． 一般如果采取的动作依赖于 Markov 链的历史，那就复杂多了． 但是在实际问题中，出现更多的是只依赖当前所处的状态的情形，或者近似是这种情形． 作为个别情形，也需要考虑依赖最近两个或多个时间的历史资料的情形，此时原则上可以通过扩大状态空间维数的方法，化为 Markov 策略的情形．

定义１６．５　（最佳策略）　Markov 随机决策的基本问题是：寻找最优平均累计报酬$\sup_{\mathbf{f}} J(m_0,\mathbf{f})$及最佳马氏策略$\mathbf{f}^*$，使平均累计报酬

$$J(m_0,\mathbf{f}^*) = \sup_{\mathbf{f}} J(m_0,\mathbf{f}), \qquad (16.5)$$

这里 Sup 取遍所有的 Markov 策略．这样的$\mathbf{f}^*$如果存在，则称为最佳 Markov 策略．

在实际问题中，有时即使最佳 Markov 策略存在，但却因为计算量过大或计算时间过长而变得实际上不可能．而e－最佳 Markov 策略常比最佳 Markov 策略更为实际可行．

定义１６．６　（e－最佳 Markov 策略）　$\mathbf{f}^e$称为一个e－最佳 Markov 策略，如果

$$J(m_0,\mathbf{f}^e) > \sup_{\mathbf{f}} J(m_0,\mathbf{f}) - e. \qquad (16.6)$$

［注］ 在求最佳 Markov 策略的计算复杂度为非多项式时，求e－最佳马氏策略常常能使计算降为多项式复杂度．可见花费这个e的代价，能带来极大的收益．

定义１６．７　如果对于任意n，a_n也不依赖于x_n，但是，是随机的，即它是与 Markov 链的发展无关的"常随机"行动：$a_n = h_n$(h_n是取值于A的随机变量)，这种策略记为 $\boldsymbol{a} = \{h_n, n \le m\}$，称为独立的随机策略．

2．2 有限时段总报酬准则下的最佳 Markov 策略的构造

定理１６．８　在所有的 Markov 策略类中存在最佳 Markov 策略(但未必唯一)．

证明　证明的过程实际上就是寻找一个最佳 Markov 策略的具体方法．要在 Markov 策略类中构造一个最佳的．而构造最佳 Markov 策略$\mathbf{f}^*$的方法则完全和例１６．１一样．

首先取$f_m^*(i)$

$$g_m(i, f_m^*(i)) = \max_{a\in A} g_m(i,a) (\overset{\Delta}{=} g_m^*(i)) .$$

再取 f_{m-1}^* 使时刻 $m-1$ 的报酬函数达到最大，即

$$g_{m-1}(i, f_{m-1}^*(i)) + \sum_{j=1}^{N} p_{ij}(f_{m-1}^*(i)) g_m^*(j) = \max_{a\in A}[g_{m-1}(i,a) + \sum_{j=1}^{N} p_{ij}(a) g_{m-1}^*(j)] (\overset{\Delta}{=} g_{m-1}^*(i))$$

然后，向后递推地取 $f_k^*(i)$ 使时刻 k 的报酬函数达到最大，即

$$g_k(i, f_k^*(i)) + \sum_{j=1}^{N} p_{ij}(f_k^*(i)) g_{k+1}^*(j) = \max_{a\in A}[g_k(i,a) + \sum_{j=1}^{N} p_{ij}(a) g_k^*(j)] (\overset{\Delta}{=} g_k^*(i)) .$$

这样就得到了 $\mathbf{f}^* = \{f_0^*, \cdots, f_{m-1}^*, f_m^*\}$． 现在证明它是最佳的 Markov 策略． 对于任意 Markov 策略 f=(f_n ,0≤n≤m)，定义辅助 Markov 策略 $\mathbf{f}^{(k)} \overset{\Delta}{=} \{f_0, \cdots, f_k, f_{k+1}^*, \cdots, f_m^*\}$．用后向数学归纳法可以直接验证以下不等式

$$J(m_0, \mathbf{f}) \le J(m_0, \mathbf{f}^{(k)}) \quad (k\le m).$$

取 k=0 即得 $J(m_0, \mathbf{f}) \le J(m_0, \mathbf{f}^*)$． 请读者自己补上这段验证．

[注 1] 以上寻找 $\mathbf{f}^*$ 的方法，在计算机上实现十分简单．问题在于当行动集 A 较大的时候，计算量会非常大，甚至难以在允许的时间内完成。于是代之以用 $e-$最佳 Markov 策略．

[注 2] 本定理可以推广到状态集 S 是为可数集，行动集 A 是紧集，而报酬函数 $g_n(i,a)$ 是有界连续函数情形．

[注 3] 在状态集与动作集较为一般的时候，为了保证最佳策略的存在，纯策略类就不够,必须考虑用随机策略类．

2．3 无穷时段下的总报酬情形（$m=\infty$ 的情形）

定理１６．９ 如果报酬函数序列满足以下的衰减性质

$$\sum_{n=0}^{\infty} \sup_{i,a} |g_n(i,a)| < \infty, \qquad (16.7)$$

那么存在最佳 Markov 策略 $\mathbf{f}^*$． （可以利用有限近似证明）．

例１６．１０ （折扣模型） 在应用中常见例子为折扣报酬模型，即

$$g_n(i,a) = r^n g_0(i,a)$$

的情形，其中 $0<r<1$ 是折扣因子．此时只要 $g_0(i,a)$ 是有界函数，则条件(16.7)满足．

定义１６．１１ （平稳策略） 如果 Markov 策略 f=(f_n , n≥0)对于任意 n 满足 $f_n = f_0$，则称为平稳 Markov 策略．

定理１６.１２ 对于折扣报酬模型，最佳 Markov 策略（如果存在）是平稳 Markov 策略.

证明 记平均累计报酬为

$$J(i, \mathbf{f}) = \sum_{n=0}^{\infty} r^n E g_0(x_n, a_n).$$

假定最佳 Markov 策略是 $\mathbf{f}^* = (f_n^*, n \geq 0)$. 令

$$\mathbf{f}^{(k)} \stackrel{\Delta}{=} \{f_0^*, \cdots, f_0^*, f_1^*, f_2^*, \cdots\} \quad (\text{共} k \text{个} f_0^*).$$

那么，用归纳法可以证明

$$J(i, \mathbf{f}^{(n)}) = J(i, \mathbf{f}^{(n-1)}) = \cdots = J(i, \mathbf{f}^{(0)}) = J(i, \mathbf{f}^*).$$

令 $n \to \infty$，便得 $J(i, \mathbf{f}^{(\infty)}) = J(i, \mathbf{f}^*)$. 这说明平稳 Markov 策略 $\mathbf{f}^{(\infty)} \stackrel{\Delta}{=} (f_0^*, f_0^*, \cdots)$ 是一个最优 Markov 策略. 】

利用不动点理论还可以证明：对折扣报酬模型，若状态集 S 可数，行动集 A 是紧集，且 $g_0(i,a)$ 有界，则平稳 Markov 策略 f 是最佳策略的充要条件是，它满足下述 Bellman 方程（动态规划方程）：

$$J(i, \mathbf{f}) = \max_{a \in A} [g_0(i,a) + r \sum_j p_{ij}(a) J(j, \mathbf{f})], \quad (\forall i \in S). \qquad (16.8)$$

由此可知 J(.,f) 是下述非线性变换 $J \to T(J)$ 的唯一不动点（即 $J(.,\mathbf{f})$ 满足方程 $T(J) = J$）：

$$T(J)(i) \stackrel{\Delta}{=} \max_{a \in A} [g_0(i,a) + r \sum_j p_{ij}(a) T(J)(j)], \quad (\forall i \in S).$$

这个结论的优点在于：定理中的 m 可用∞代替（即近似），后者显然更为简单，因为它的最佳策略是平稳 Markov 策略，而且只要求出变换 T 的不动点（即 $T(J) = J$ 的解 $J(.)$），再令

$$f(i) \stackrel{\Delta}{=} a^*$$

即可，其中 a^* 是 $[g_0(i,a) + r \sum_j p_{ij}(a) T(J)(j)]$ 达到最大值的动作. 这样的计算将大大地减少机耗时间. 有时对于充分大的 m，可以用 $T_m(I)$ 来近似 $J(.)$，此处 I 是恒同映射，而 $T_m \stackrel{\Delta}{=} T \circ \cdots \circ T$ 为映射 T 的 m 次复合.

[注 5] 以上是以平均累计报酬作为目标函数的. 它是按期望准则取最优，还可以按矩准则取最优，按样本路径平均取最优，后者的报酬是一个随机变量：记时刻 m 前的累计报酬为 $V_m(i_0, \mathbf{f})$，其中 i_0 为系

统的初始状态．作平均报酬序列 $J_N(i_0,\mathbf{f})\overset{\Delta}{=}\frac{1}{N+1}\sum_{m=0}^{N} V_m(i_0,\mathbf{f})$ ，一般 $J_N(i_0,\mathbf{f})$ 的极限未必存在，所以常取其收敛子序列中极限的最大者， 记为 $\overline{J}(i_0,\mathbf{f})$ ，或极限中的最小者， 记为 $\underline{J}(i_0,\mathbf{f})$ ．用它们作为目标函数． 前者是以乐观的态度看待报酬，而后者是以悲观的态度看待报酬．在使用随机策略时， 相应地记为 $\underline{J}(i_0, p)$ ， 其中 $p=(p_0,p_1,\cdots),p_n$ 表示时刻 n 使用的随机策略． 而把 $\underline{J}(i_0)\overset{\Delta}{=}\sup_p \underline{J}(i_0,p)$ 作为最佳函数． 这种平均准则常见于发展较为平稳的系统，例如通讯网络，而累计总报酬准则，则常见于较快变化的系统，例如金融系统．轨道平均报酬准则的数学处理，通常比平均累计总报酬准则情形要复杂．

［注６］ 最一般的模型是策略的动作依赖历史情况的策略：设 S 与 A 仍都为有限集，又设随机变量列 $\{x_n\}$ 满足

$$P(x_{n+1}=j\,|\,x_0=i_0,x_1=i_1,\cdots,x_n=i_n)=p_{i_n,j}(a_n)\,,$$

其中 a_n 由下式递推地确定：

$$a_0=f_0(i_0)\,,$$

$$\cdots,$$

$$a_n=f_n(i_0,a_0,i_1,a_1,\cdots,i_{n-1},a_{n-1},i_n)\,.$$

这时的 $\{x_n\}$ 不再是 Markov 链，因为它依赖于所有的过去历史． 而且此时 a_n 是一个依赖于 $\{x_0,\cdots,x_n\}$ 的随机变量．如果我们仍记 f=(f_n, n≥0)，并称之为一个策略，那么对此模型也可类似地定义最佳策略及最佳 e －策略．事实上，对此模型可以证明：在很宽的条件下，定理１６．８中的最佳 Markov 策略也是这里的策略类中的最佳策略．也就是最佳 Markov 策略在以上定义的非马氏策略类中仍是最佳的．这就可以把搜索最佳策略的策略类的范围大大地缩小了．

［注７］ 在常见的实际模型中，有一类状态空间 S 是一个区间(端点可以为 ∞)，动作集 $A=A_x$ 是依赖状态 x 的一个有限区间．例如

(1) Ricker 模型（合理捕鱼问题） 设在时刻 n 湖中鱼的总量为 x_n ．捕鱼的原则是，在有计划地留下数量(随机的) a_n 的鱼用以繁殖的前题下，求最大捕鱼量． 假定鱼的数目与留下的数量受群体控制的 Ricker 模型所制约：

$$x_{n+1}=c_1a_ne^{h_n-c_2a_n}\,,$$

其中 c_1,c_2 为常数， $\{h_n\}$ 是独立同分布随机变量列． 于是捕鱼量为 $g\ (x_n,a_n)\overset{\Delta}{=}x_n-a_n$ 中的 $g_n(x,a)=x-a$ 与 n 无关．这时状态空间为 $S=[0,\infty)$ ，当状态确定为 x 时的动作集为

$A_x=[0,x]$． 在这模型中，相应于转移矩阵 P $(a)=(p_{ij}(a),a\in A_i)$ 的是转移核 $p(x,y;a)$

$(a\in A_x)$．如果采用策略 $\mathbf{f}=(f_n,a_n=f_n(x_n))$，则在时刻 m 的平均累计报酬为

$J(x_0,\mathbf{f})=E_{x_0}[\sum_{n=0}^{m}(x_n-f(x_n))]$，其中 $x_0=x_0$ (也可以是随机的)，$0\le f_n(x)\le x$．

（2）最佳投资组合 假定只有两种证券，无风险的证券(例如存银行，其利率为 r)及有风险的股票． 设财产 x_n 中投资于股票的比例为 $a_n,(0\le a_n\le 1)$，且以数量 c_n 消费． 此时状态空间为 $S=[0,\infty)$，当状态确定为 x 时的动作集为 $A_x=[0,1]\times[0,x]$，即

$$a=(a,c),(0\le a\le 1,0\le c\le x).$$

于是状态发展为

$$x_{n+1}=[(1+r)(1-a_n)+a_n h_n](x_n-c_n),$$

其中 h_n 是股票的价值．典型的报酬函数是 $g(x,(a,c))=u(c)$，它代表由消费带来的 "乐趣"，在经济学中称为由消费 c 产生的效用函数． 如果假定 $\{h_n\}$ 为独立同分布，则 $\{x_n\}$ 正是与前面类似的模型．但是，用独立同分布的随机变量来描述股票的误差较大．如果用 Black-Scholes 模型的离散采样描述股票，虽然较为合理，但是这会使模型变得复杂．

（3）存储问题 讨论时刻 n 时商店中某种货品的存储量 x_n．设进货量为 a_n，假定市场需求为独立同分布的随机变量序列 h_n．又设商店对此货品划定了最大容许库存量 c．那么，状态空间 $S=[-\infty,c]$ (在 $x_n<0$ 时表示缺货)，在状态(库存)为 x 时的动作集为 $A_x=[0,c-x]$，而

$$x_{n+1}=x_n+a_n-h_n.$$

此时报酬函数依赖于市场需求． 假定单个货品的卖价为 s，进货价为 d，存储消耗价为 h，那么时刻 n 的报酬为

$$g_n(x_n,a_n,h_n)=s(\min(x_n+a_n,h_n)-da_n-h(x_n+a_n).$$

由此可以考虑最大化报酬．

［注８］还可以考虑时间连续的情形，这就是可控 Markov 过程． 这类问题较为复杂，在概率论中涉及最佳停时，而在求解最佳策略时，常用偏微分方程中的活动边界理论或变分不等式．

龚光鲁，钱敏平著　应用随机过程教程 – 与在算法和智能计算中的应用
清华大学出版社，2003

第 17 章　Poisson 随机分析简介与典型的点过程

1. 非时齐的 Poisson 过程与非时齐的复合 Poisson 过程与特征泛函

1．1　数值函数对 Poisson 过程的积分

定义 17．1　设 N_t 是一个强度为 l 的 Poisson 过程，对应的更新流为 $\{t_n\}$，

$T_n = t_n - t_{n-1} \sim Exp_l$．　定义 $[0,\infty)$ 上的函数 $f(t)$ 关于 N_t 的积分为

$$\int_0^t f(s)dN_s = \begin{cases} \sum_{n=1}^{N_t} f(t_n) & (N_t \ge 1)) \\ 0 & (N_t = 0) \end{cases}. \qquad (17.\ 1)$$

在 t 给定时，它是一个随机变量，其含义为 $f(s)$ 在到时刻 t 为止的指数流上的函数值之和．如果把 $f(s)$ 看成在时刻 s 发生事故的代价，那么这个积分就表示到时刻 t 为止，由指数流描述的事故流所付出的总代价．由定义显见有 $\int_0^t \mathbf{1}\, dN_s = N_t$．

1．2　Poisson 过程的特征泛函

定义 17.2　对于 Poisson 过程 N_t 及定义在 $[0,T]$ 上的函数 $f(t)$，我们把 $\int_0^T f(s)dN_s$ 的特征函数在 1 处的值记为 $\Phi_N(f)$，即

$$\Phi_N(f) = Ee^{i\int_0^T f(s)dN_s}.$$

于是对于给定一个函数 f，就有一个数 $\Phi_N(f)$ 与之对应，这种从函数 f 到 $F(f)$ 的映射称为泛函，又因为此泛函是通过 Poisson 过程的积分生成的，所以称为 Poisson 过程的特征泛函．

例 17．3　当 $f(s) \equiv J \cdot I_{[0,t]}(s)$ 时，Poisson 过程的特征泛函就简化为 Poisson 过程在时刻 t 的特征函数 $Ee^{iq \cdot N_t}$．　而当 $f(t) \equiv J_1 \cdot I_{[0,t_1]}(t) + J_2 \cdot I_{(0t_2]}(t)$ 时，特征泛函就简化为 Poisson 过程在时刻 t_1 与时刻 t_2 的联合特征函数 $Ee^{i(q_1 \cdot N_{t_1} + J_2 N_{t_2})}$．

设 $t_1 < t_2 \le T$．那么利用 Poisson 过程的独立增量性与时齐性，对于

$f(t)=J_1\cdot I_{[0,t_1]}(t)+J_2\cdot I_{(t_1,t_2]}(t)$，我们得到

$$\Phi(f)=F(J_1\cdot I_{[0,t_1]}+J_2\cdot I_{(t_1,t_2]})=Ee^{i(q_1\cdot N_{t_1}+q_2(N_{t_2}-N_{t_1}))}=Ee^{iq_1\cdot N_{t_1}}Ee^{iq_2 N_{t_2-t_2}}$$

$$=e^{lt_1(e^{iJ_1}-1)}\,e^{l(t_2-t_1)(e^{iJ_2}-1)}、=e^{\int_0^T l(e^{i[q_1 I_{[0,t_1]}(t)+q_2 I_{(t_1,t_2]}(t)]}-1)dt}=e^{\int_0^T l(e^{if(t)}-1)dt}.$$

再复杂一些，对于$0=t_0<t_1<\cdots<t_n\le T$，及$f(t)=\sum_{k=0}^{n-1}c_k I_{(t_k,t_{k+1}]}(t)$，我们类似地可以

得到

$$Ee^{i\int_0^T f(t)dN_t}=\Phi_N(f)=e^{\int_0^T l(e^{if(t)}-1)dt}.$$

这个公式对于任意连续函数，甚至更为一般的函数，可以利用函数逼近的方法证明它仍然正确．于是我们得到如下的定理．

定理 17．4　Poisson 过程的特征泛函的表达公式为

$$Ee^{i\int_0^T f(t)dN_t}=\Phi_N(f)=e^{\int_0^T l(e^{if(t)}-1)dt}. \qquad (17．2)$$

此定理是 Poisson 随机变量的特征函数的自然推广．

１．３　非时齐 Poisson 过程的统计性质

我们回忆非时齐的 Poisson 过程N_t，它是非时齐的独立增量过程．它与时齐的 Poisson 过程的不同之处仅仅在于其强度不再是一个常数l，而是一个依赖于时间的函数$l(t)$，称为非时齐的 Poisson 过程的强度函数．即对于$s<t$，随机增量N_t-N_s 服从参数为$\int_s^t l(u)du$的 Poisson 分布．据此可以通过 Poisson 分布，对非时齐的 Poisson 过程作随机模拟．

设N_t是一个强度为$l(t)$的非时齐的 Poisson 过程，它是一种记录”事故”的计数过程，将这个过程的各个计数(事故)发生的随机时刻记为 $0=t_0<t_1<\cdots<t_n<\cdots$．它们间仍由等式

$$\{N_t\ge n\}=\{t_n\le t\}$$

相联系．与时齐的 Poisson 过程相比，$\{t_n\}$不再是更新流（稍后将证明它是 Markov 链）．

定理 17．5　前n个发生时刻$(t_1,\cdots,t_n)$的联合分布密度$g_{t_1,\cdots,t_n}(s_1,\cdots,s_n)$的表达式为

$$g_{t_1,\cdots,t_n}(s_1,\cdots,s_n)=l(s_1)\cdots l(s_n)e^{-\int_0^{s_n} l(u)du} I_{\{0\le s_1<\cdots<s_n\}}\,. \qquad (17.3)$$

证明 仿照时齐的 Poisson 过程的情形便得.

推论 17.6 发生时刻列$\{t_n\}$是状态连续的非时齐的 Markov 链.

证明 在已知$(t_1,\cdots,t_n)=(s_1,\cdots,s_n)$的条件下，随机变量$t_{n+1}$的条件分布密度为

$$g_{t_{n+1}|t_1,\cdots,t_n}(s_{n+1}\mid s_1,\cdots,s_n)=l(s_{n+1})e^{\int_{s_n}^{s_{n+1}} l(u)du} I_{s_n<s_{n+1}}\,,$$

它只与t_n的取值s_n有关.

推论 17.7 在已知$(t_1,\cdots,t_n)=(s_1,\cdots,s_n)$的条件下，第$n+1$个随机间隔$T_{n+1}=t_{n+1}-t_n$的条件分布密度为

$$g_{T_{n+1}|t_1,\cdots,t_n}(t\mid s_1,\cdots,s_n)=l(s_{n+1})e^{\int_{s_n}^{s_n-t} l(u)du} I_{s_n<s_{n+1}}\,. \qquad (17.4)$$

定理 17.8 时刻t时的计数N_t与发生时刻的联合分布（注意这是混合型的随机向量）$P(N_t=n,t_1\le s_1,\cdots,t_n\le s_n)$关于$(s_1,\cdots,s_n)$的密度（称为 Poisson 过程的样本分布）为

$$p_{N_t,t_1,\cdots,t_{N_t}}(n,s_1,\cdots,s_n)=[\,l(s_1)\cdots l(s_n)I_{\{0\le s_1<\cdots<s_n,n>0\}}+I_{\{n=0\}}]e^{-\int_0^t l(u)du}\,. \quad (17.5)$$

证明 利用条件概率

$$P(N_t=n\mid t_1=s_1,\cdots,t_n=s_n)=P(N_t-N_{s_n}=0)=e^{-\int_{s_n}^t l(u)du}$$

及定理 17.5 即得. 】

非时齐的 Poisson 过程的时齐随机分流定理仍然成立，即如果将强度函数为$l(t)$的非时齐的 Poisson 过程N_t的发生的各个事故以概率p和$1-p$与N_t独立地分别归入第 1 类和第 2 类，那么，第 1 类发生时刻列是一个强度函数为$p\,l(t)$的非时齐 Poisson 过程$N_t^{(1)}$的事件发生时刻列.

同样，非时齐的 Poisson 过程的非时齐分流定理也是成立的.

定理 17.9 设$h_1,\cdots,h_n$是独立同分布的随机变量，其分布密度为

$$\frac{l(t)}{\int_0^t l(u)du} I_{[0,\infty)}(t) .\qquad (17.6)$$

那么，在$N_t = n$的条件下，发生时刻$(t_1, \cdots, t_n)$的条件分布与$(h_1^*, \cdots, h_n^*)$同分布，其中$h_1^*, \cdots, h_n^*$是$h_1, \cdots, h_n$按由小到大排列后的次序随机变量列：$h_1^* < \cdots < h_n^*$.

证明　仿照时齐的 Poisson 过程的相应定理的证明.　　　　　　　　】

定理１７.９可以用以对给定强度的非时齐的 Poisson 过程作计算机模拟.

例１７．１０　放射元发射的光子数目可以用强度为

$$l(t, a, b) = a \cdot e^{-b \cdot t} \qquad (a, b > 0)$$

的非时齐的 Poisson 过程来建模．它模拟了放射强度的衰变情况．而在核医疗中使用的光脉冲序列通常用强度为

$$l(t, g, a_1, \cdots, a_n, b_1, \cdots, b_n) = g + \sum_{k=1}^{n} a_i e^{-b_i t} \qquad (a_k, b_k, g > 0)$$

的非时齐的 Poisson 过程来模拟并建模．而调幅，调相与调频为f_m的脉冲光源则可以分别用强度为

$$l(t, g, a, b) = g + ab \,|\, S(t) \,|^2,$$

$$l(t, g, a, b) = g + a \,|\, S(t - b) \,|^2$$

与

$$l(t, g, a, b, m) = g + a(1 + m\cos[2p(f_m + b)t]) \qquad (|m| < 1)$$

的非时齐的 Poisson 过程模拟.

１．４　数值函数对非时齐 Poisson 过程的积分及非时齐的 Poisson 过程的特征泛函

我们仍可以定义$[0, \infty)$上的数值函数$f(t)$关于非时齐 Poisson 过程N_t的积分为

$$\int_0^t f(s)dN_s = \begin{cases} \sum_{n=1}^{N_t} f(t_n) & (N_t \ge 1) \\ 0 & (N_t = 0) \end{cases} .\qquad (17.7)$$

同样，在t给定时它是一个随机变量．如果把$f(s)$看成在时刻s发生事故所付出的代价，那么这个随机积分仍表达到时刻t为止，由此非时齐的 Poisson 过程所描述的事故流所付出的总代价.

定义 17．１１　非时齐的 Poisson 过程的特征泛函仍定义为

$$\Phi_N(f) = Ee^{i\int_0^T f(t)dN_t} .$$

特征泛函是有限个时刻的联合特征函数在一个区间上的所有时刻情形的自然推广． 由于在 $s<t$ 时有

$$Ee^{iq(N_t-N_s)}=e^{\int_s^t l(u)du(e^{iq}-1)},$$

于是对于 $f(t)=J_1\cdot I_{[0,t_1]}(t)+J_2\cdot I_{(t_1,t_2]}(t)$**就有**

$$\Phi(f)=\Phi(J_1\cdot I_{[0,t_1]}+J_2\cdot I_{(t_1,t_2]})=Ee^{i(q_1\cdot N_{t_1}+q_2(N_{t_2}-N_{t_1}))}=Ee^{iq_1\cdot N_{t_1}}Ee^{iq_2(N_{t_2}-N_{t_1})}$$

$$=e^{\int_0^{t_1} l(u)du(e^{iJ_1}-1)}e^{\int_{t_1}^{t_2} l(u)du(e^{iJ_2}-1)}`=e^{\int_0^T l(t)(e^{i(q_1I_{[0,t_1]}(t)+q_2I_{(t_1,t_2]}(t))}-1)dt}=e^{\int_0^T l(t)(e^{if(t)}-1)dt}.$$

我们可以归纳为如下的定理

定理１７．１２　对于连续函数 $f(t)$ (**或更为一般的有界** Borel **函数** f)，**非时齐的** Poisson **过程的特征泛函在** f **处的值为**

$$Ee^{i\int_0^T f(t)dN_t}=\Phi_N(f)=e^{\int_0^T l(t)(e^{if(t)}-1)dt}.\qquad\text{（１７．８）}$$

在一些统计问题中常常用到在强度函数 $l(t)$ **中带有未知参数的非时齐的** Poisson **过程**（**参见例** 17.10)，**这时需要用实测数据来估计这些未知参数，例如用最大似然估计，于是需将观测到的样本置入定理** 17．**８的分布中去，为此，我们要对定理１７．８作如下的变形**

定理 17．８’　对于非时齐的 Poisson 过程，我们有(由过程的一个样本给出的似然函数)：在 $N_t=0$ **时**

$$p_{N_t,t_1,\cdots,t_{N_t}}(N_t,t_1,\cdots,t_n)=e^{-\int_0^t l(u)du},$$

而在 $N_t>0$ **时**

$$p_{N_t,t_1,\cdots,t_{N_t}}(N_t,t_1,\cdots,t_n)=e^{-\int_0^t l(u)du+\sum_{k=1}^{N_t}\ln l(t_k)}=e^{-\int_0^t l(u)du+\int_0^t \ln l(u)dN_u}.\qquad\text{（１７．９）}$$

［注］　在用例１７.１０的参数模型建模时，需要用过程的一段观测轨道估计未知参数．而使（１７．９）取得最大值的参数作为估计，就是最大似然估计．

推论 １７．１３

$$E(\int_0^t f(s)dN_s)=\int_0^t f(s)l(s)ds,\qquad\text{（１７．１０）}$$

$$Var(\int_0^t f(s)dN_s)=\int_0^t f(s)^2l(s)ds.\qquad\text{（１７．１１）}$$

证明　由定理１７．９，利用对称性，我们得到

$$\sum_{k=1}^{n} E[f(t_k) \mid N_t = n] = E(\sum_{k=1}^{n} f(h_k^*)) = E(\sum_{k=1}^{n} f(h_k)) = \sum_{k=1}^{n} Ef(h_k)$$

$$= n\int_0^t f(u)\frac{l(u)}{\int_0^t l(s)ds}du . \qquad （１７．１２）$$

于是

$$E(\int_0^t f(s)dN_s) = \sum_{n=0}^{\infty} \sum_{k=1}^{n} E[f(t_k) \mid N_t = n]P(N_t = n)$$

$$= EN_t\int_0^t f(u)\frac{l(u)}{\int_0^t l(s)ds}du = \int_0^t f(u)\,l(u)du .$$

此即（１７．１０）.（１７．１１）的证明是类似的.

注　也可以用特征泛函$\Phi_N(J \cdot f)$对J求一阶微商和二阶微商得到.

命题１７．１４　（非时齐的 Poisson 过程的补偿函数）　设N_t是以$l(t)$为强度函数的非时齐的 Poisson 过程，那么$\tilde{N}_t \stackrel{\Delta}{=} N_t - \int_0^t l(s)ds$是鞅．$\Lambda_t \stackrel{\Delta}{=} \int_0^t l(s)ds$称为非时齐的 Poisson 过程的补偿函数.

定义１７．１５　（对非时齐的 Poisson 过程的随机积分）　对于有界的(N_t)可知的随机过程Ψ_t，用与 Ito 积分类似地用积分和的极限，可以定义Ψ_t关于鞅$\tilde{N}_t$的随机积分，以及关于N_t的随机积分：

$$\int_0^t \Psi_s dN_s = \begin{cases} \sum_{n=1}^{N_t} \Psi_{t_n} & (N_t \geq 1)) \\ 0 & (N_t = 0) \end{cases} .$$

关于鞅$\tilde{N}_t$的随机积分是 Ito 积分的非时齐的 Poisson 版本，而且有

$$\int_0^t \Psi_s dN_s = \int_0^t \Psi_s d\tilde{N}_s + \int_0^t \Psi_s\, l(s)ds$$

关于时齐的 Poisson 过程的随机积分有许多与 Ito 积分相仿的性质.

命题１７．１６　鞅$\tilde{N}_t$的特征泛函为

$$\Phi_{\tilde{N}}(f) \stackrel{\Delta}{=} Ee^{i\int_0^T f(s)d\tilde{N}_s} = e^{\int_0^T [e^{if(s)}-1-if(s)]l(s)ds} \quad .（１７．１３）$$

推论１７．１７

（１）　$E(\tilde{N}_s\tilde{N}_t) = Cov(N_s, N_t) = \int_0^{s\wedge t} l(u)du$ ．

（２）　$E(\tilde{N}_{t_1}\tilde{N}_{t_2}\tilde{N}_{t_3}) = \int_0^{t_1\wedge t_2\wedge t_3} l(u)du$ ．

（３）　$E(\tilde{N}_{t_1}\tilde{N}_{t_2}\tilde{N}_{t_3}\tilde{N}_{t_4}) = \int_0^{t_1\wedge t_2\wedge t_3\wedge t_4} l(u)du + Cov(N_{t_1}, N_{t_2})\,Cov(N_{t_3}, N_{t_4}) +$

$+Cov(N_{t_1}, N_{t_3})\,Cov(N_{t_2}, N_{t_4}) + Cov(N_{t_1}, N_{t_4})\,Cov(N_{t_2}, N_{t_3})$ ．

证明　取 $f(t) = \sum_{k=1}^{4} J_i I_{[0,t_k]}(t)$ ．将 $\Phi_{\tilde{N}}(f)$ 记为 $j(J_1,\cdots,J_4)$ ．求 $\frac{\partial^4 j}{\partial J_1\cdots\partial J_4}(0,0,0,0,)$，就得到（３）．其它类似．

［注］　对于一般未必可微的递增函数 Λ_t，还可以推广定义以 Λ_t 为补偿函数的非时齐的 Poisson 过程 N_t： 非时齐的独立增量过程， 且对于任意 $s<t$，有 $N_t - N_s \sim Poisson_{\Lambda_t-\Lambda_s}$． 此时仍有与（１７．８),（１７．１０),（１７．１１）与（１７．１２）相应的结论．

１．５　非时齐的复合 Poisson 过程及其特征泛函

定义１７.１８　设 N_t 为非时齐的 Poisson 过程，$\{X_n\}$ 为与之独立的独立同分布随机变量序列．$Y_t = \sum_{k=1}^{N_t} X_k$ 称为非时齐的复合 Poisson 过程．$\{X_n\}$ 称为赋值随机变量序列（或标值序列)．

可以证明非时齐的复合 Poisson 过程是非时齐的独立增量过程．

命题１７.１９　非时齐的复合 Poisson 过程 Y_t 在区间 $[0,T]$ 上的特征泛函定义为

$$\Phi_Y(f) \stackrel{\Delta}{=} Ee^{i\int_0^T f(t)dY_t} \quad .\qquad（１７．１４）$$

于是有

$$\Phi_Y(f) = e^{\int_0^T [j(f(u)-1])\,l(u)du} \quad ,\qquad（１７．１５）$$

其中 $j(J)$ 是 X_1 的特征函数：$j(J)=Ee^{iJX_1}$．

证明　与（１７．１２）类似地有

$$E[e^{i\sum_{k=1}^{N_T} f(t_k)X_k} \mid N_T=n)=Ee^{i\sum_{k=1}^{n} f(h_k)X_k}=E[\prod_{k=1}^{n} j(f(h_k))]$$

$$=\frac{1}{(\int_0^T l(u)du)^n}[\int_0^T j(f(u))\,l(u)du]^n .$$

于是由全期望公式得到

$$\Phi_Y(f)=Ee^{iI_{\{k\ge 1\}}\sum_{k==1}^{N_T} f(t_k)X_k}$$

$$=P(N_T=0)+\sum_{n=1}^{\infty} E(e^{i\sum_{k=1}^{n} f(t_k)X_k} \mid N_T=n)P(N_T=n)$$

$$=e^{-\int_0^T l(u)du}[1+\sum_{n=1}^{\infty}\frac{1}{(\int_0^T l(u)du)^n}[\int_0^T j(f(u))\,l(u)du]^n\frac{(\int_0^T l(u)du)^n}{n!}]$$

$$=e^{-\int_0^T l(u)du}e^{\int_0^T j(f(u))\,l(u)du}=e^{\int_0^T[j(f(u)-1])\,l(u)du} .$$ 】

因为具有相同的特征泛函的两个随机过程的统计性质是一样的，所以命题１７．１９常用于由特征泛函的形式来确定一个随机过程是否是非时齐的复合 Poisson 过程．

例１７．２０　（非时齐的广义 Poisson 过程）

如果赋值随机变量具有离散分布 $X_n\sim\begin{pmatrix}1 & \cdots & k & \cdots\\ p_1 & \cdots & p_k & \cdots\end{pmatrix}$，则非时齐的复合 Poisson 过程称为非时齐的广义 Poisson 过程．此时的特征泛函为

$$\Phi_Y(f)=e^{\int_0^T [\sum_{k=1}^{\infty} p_k e^{if(u)k}-1])\,l(u)du} . \quad (１７．１６)$$

２．与非时齐的复合 Poisson 过程相系的 Poisson 点过程

２.１　将非时齐复合 Poisson 过程表为非时齐 Poisson 过程的积分(用时间积分表示)

利用对于非时齐的 Poisson 过程的随机积分的定义，立得下述命题．

命题１７．２１　（非时齐的复合 Poisson 过程的非时齐 Poisson 积分表示）

设 N_t 为非时齐的 Poisson 过程，$\{X_n\}$ 为与之独立的独立同分布随机变量序列．而

$Y_t=\sum_{k=1}^{N_t} X_k$是非时齐的复合 Poisson 过程，则它有以下的积分表示式

$$Y_t=\int_0^t \tilde{X}_s\, dN_s\text{，} \qquad (17.17)$$

其中 $\tilde{X}_s=\begin{cases}X_n & (s=t_n)\\ 0 & (\text{其它})\end{cases}$．于是非时齐的复合 Poisson 过程的特征泛函也可以写为

$$Ee^{i\int_0^T \tilde{X}_t dN_t}=e^{\int_0^T [j(f(u)-1])l(u)du}. \qquad (17.15)'$$

2．2　将非时齐复合 Poisson 过程表为 Poisson 点过程的积分（用空间积分表示）

我们先讨论简单的情形．对于离散赋值的非时齐的复合 Poisson 过程$Y_t=\sum_{k=1}^{N_t} X_k$．

设 $X_n\sim\begin{pmatrix}v_1 & \cdots & v_k & \cdots\\ p_1 & \cdots & p_k & \cdots\end{pmatrix}$．将过程在时刻$t$以前取值$v$的累计次数记为

$$N_t(\{v\})=\text{集合}\{Y_s=v:s\le t\}\text{中的元素个数}. \qquad (17.18)$$

对于固定的值v，$N_t(\{v\})$可以看成对于非时齐的 Poisson 过程 N_t 的分流．由随机分流定理可知，$N_t(\{v_k\})$是强度函数为 $p_k\,l(t)$ 的非时齐 Poisson 过程．

由非时齐的复合 Poisson 过程的定义，直观地可以看出，对于$v=v_1,\cdots,v_k,\cdots$，$\{N_t(\{v\})\}$是一系列相互独立的非时齐的 Poisson 过程．而且有

$$Y_t=\sum_k v_k N_t(\{v_k\}). \qquad (17.19)$$

又因为 Poisson 过程序列$\{N_t(\{v\})\}$（$v=v_1,\cdots,v_k,\cdots$）与空间的取值点列$\{v_1,\cdots,v_k,\cdots\}$相系，所以称$\{N_t(\{v\})\}$为 Poisson 点过程．Poisson 点过程的概念要比 Poisson 过程更复杂，它包含时空两个参数(t,v)，而且对于不同的v，作为时间t的随机过程是相互独立的．

表达式（17．19）的直观含义是，如果把非时齐的复合 Poisson 过程Y_t看成在时刻t的总的随机积累，那么它是时刻t前取大小不同的固定值的随机积累的总和．

这个思想可以拓广到一般的非时齐的复合 Poisson 过程Y_t，即随机变量 X_k 不必局限于取离散值，而是可以取任意值(甚至可取负值)的情形．这时Y_t也可以取任何的值．它在时刻t前取值于区间$(a,b]$的次数，是一个随机过程，记之为 $N_t((a,b])$（因为Y_t当且仅当在

此非时齐的 Poisson 过程N_t的事件列t_n上跳跃，所以这个次数是一个有限的（但是随机的）数）．它们满足

（P．1）由随机分流定理，$N_t((a,b])$是强度为$P(X\in(a,b])\,l(t)$的非时齐的 Poisson 过程．

（P．2）由随机分流定理，若区间$(a_i,b_i]\quad(i=1,\cdots,m)$两两补交，则随机过程$N_t((a_i,b_i])\quad(i=1,\cdots,m)$是相互独立的．

（P．3）　对于$a<b<c$有可加性：　$N_t((a,b])+N_t((b,c])=N_t((a,c])$．

再记

$$N_t(v)\overset{\Delta}{=}N_t((-\infty,v]).\qquad（１７．２０）$$

注意　这里的记号$N_t(v)$与前面定义的记号$N_t(\{v\})$的含义是不同的．

我们将它叙述为如下的略广一些（不仅仅限于非时齐的复合 Poisson 过程）的定义．

定义１７．２２　依赖于实数值v的,正整值随机过程族$\{N_t(v):t\ge 0,-\infty<v<\infty\}$，称为 Poisson 点过程，如果对于$N_t((a,b])\overset{\Delta}{=}N_t(b)-N_t(a)$满足以上的条件（P.2），(P.3)，以及如下的(P.1)’：

（P.1）'存在单调递增函数$F(v)$，使$F(-\infty)=0$且$N_t((a,b])$是强度为$[F(b)-F(a)]\,l(t)$的非时齐的 Poisson 过程．

此时$F(v)\,l(t)$称为 Poisson 点过程的补偿函数．

例１７．２３　由非时齐的复合 Poisson 过程所定义的随机过程族$\{N_t(v):t\ge 0,-\infty<v<\infty\}$，是 Poisson 点过程，其中$N_t(v)$表示此非时齐的复合 Poisson 过程在时刻$t$前取值于区间$(-\infty,v]$的次数．

易见此 Poisson 点过程的补偿函数是$F_X(v)\,l(t)$，其中F_X是X_n的分布函数．即$\tilde{N}_t(v)=N_t(v)-F_X(v)\,l(t)$是$(N_t)$鞅．

一般地，若 Poisson 点过程的补偿函数$F(v)\,l(t)$满足$F(\infty)-F(-\infty)=C<\infty$，则存在以$C\,l(t)$为强度函数的非时齐的 Poisson 过程,及对应于赋值随机变量X_n的分布函数为$F_{X_n}(v)\overset{def}{=}\dfrac{1}{C}F(v)$的非时齐的复合 Poisson 过程，使此 Poisson 点过程由此非时齐的复合

Poisson 过程生成．

特别地，对于非时齐的 Poisson 过程N_t，则有

$$N_t(\{k\}) = I_{\{t_k \le t < t_{k+1}\}}, N_t(v) = \sum_{k \ge 1} N_t(\{k\}) I_{(k \le v < k+1\}} .$$

直观地，非时齐的复合 Poisson 过程Y_t看成在时刻t的总的随机积累，是大小不同的随机积累的总和，即对于如下的分点：$\cdots < v_i^{(n)} < v_{i+1}^{(n)} < \cdots < v_{m_n}^{(n)}, v_0^{(n)} = -n, v_{m_n}^{(n)} = n$，$\max_i (v_{i+1}^{(n)} - v_i^{(n)}) \to 0 (n \to \infty)$，$v_i^{(n)} < v_i^{(n)}{}' \le v_{i+1}^{(n)}$，有

$$Y_t \approx \sum_i v_i^{(n)}{}' N_t((v_i^{(n)}, v_{i+1}^{(n)}]) = \sum_i v_i^{(n)}{}' [N_t(v_{i+1}^{(n)}) - N_t(v_i^{(n)})] .$$

当$n \to \infty$时，可以证明上式右方按概率收敛．将此概率收敛的极限随机变量记为 $\int vN_t(dv)$．　于是就得到下面的命题．

命题１７.２４　（用 Poisson 点过程表达非时齐复合 Poisson 过程Y_t的分解式）

$$Y_t = \int vN_t(dv) . \qquad （１７．２１）$$

２．３　将非时齐复合 Poisson 过程表为时空 Poisson 过程的积分（用时空积分表示）

对于非时齐的复合 Poisson 过程Y_t．还可以从另一种角度考虑．改记

$$m(t,v) = N_t(v)$$

这是一个时空双参数的正整值随机变量族，简称为时空过程．对于$0 \le s < t, a < b$，定义

$$m((s,t] \times (a,b]) \overset{\Delta}{=} m(t,b) - m(t,a) - m(s,b) + m(s,a) , \qquad （１７．２２）$$

容易验证：

（S．１）Poisson 性：　$m((s,t] \times (a,b]) \sim Poisson_{(l(t)-l(s))(F_X(b)-F_X(a))}$，

其中$F_X(x)$是此非时齐复合 Poisson 过程相系的赋值随机变量X_k们的公共分布函数．

（S．２）独立性：若 $A_i = (s_i, t_i] \times (a_i, b_i]\ (i = 1, \cdots, m)$，且$A_i$们两两不相交，则 $m(A_1), \cdots, m(A_m)$彼此独立．

（S．３）可加性：　若 $A_i = (s_i, t_i] \times (a_i, b_i]\ (i = 1, \cdots, 3)$，且$A_i$们两两不相交，$A_1 = A_2 \bigcup A_3$，则

$$m(A_1) = m(A_2) + m(A_3) .$$

我们也给出一个一般的（不仅仅限于非时齐的复合 Poisson 过程）定义

定义１７．２５ 正整值随机过程族$\{m(t,v): t\ge 0, -\infty<v<\infty\}$，称为时空 Poisson 点过程，如果条件（S．１）',(S.2),（S．3）满足，其中$m((s,t]\times(a,b])$由（７.２３）定义，且:

（S．１）' Poisson 性： $m((s,t]\times(a,b]) \sim Poisson_{(l(t)-l(s))(F(b)-F(a))}$， F 单调递增，$F(-\infty)=0$．

注 这里只要求F 递增，并未假定 $F(\infty)<\infty$．

例１７．２６ 由非时齐的复合 Poisson 过程所定义的随机过程族

$$\{m(t,v): t\ge 0, -\infty<v<\infty\}$$

是时空 Poisson 点过程，其中$m(t,v)=N_t(v)$表示此非时齐的复合 Poisson 过程在时刻t前取值于区间$(-\infty,v]$的次数.

类似地，若时空 Poisson 点过程的补偿函数$F(v)\,l(t)$满足$F(\infty)-F(-\infty)=C<\infty$，则存在以$C\,l(t)$为强度函数的非时齐的 Poisson 过程，及对应于赋值随机变量X_n的分布函数为$F_{X_n}(v)=\dfrac{1}{C}F(v)$的非时齐的复合 Poisson 过程，使此时空 Poisson 点过程是由此非时齐的复合 Poisson 过程生成.

特别地，对于非时齐的 Poisson 过程N_t，则有

$$m((s,t]\times(a,b])=\sum_{a<k\le b} I_{\{s<t_k\le t<t_{k+1}\}}\,,$$

$$N_t=m((-\infty,t]\times(0,\infty))=\sum_k I_{\{t_k\le t\}}=\sup\{k: t_k\le t\}.$$

对于如下的分点：$\cdots<v_i^{(n)}<v_{i+1}^{(n)}<\cdots<v_{m_n}^{(n)}, v_0^{(n)}=-n, v_{m_n}^{(n)}=n$，

$0=s_0^{(n)}<\cdots<s_j^{(n)}<s_{j+1}^{(n)}<\cdots<s_n^{(n)}=t$， $\max_{i,j}(v_{i+1}^{(n)}-v_i^{(n)}, s_{j+1}^{(n)}-s_j^{(n)})\to 0(n\to\infty)$，

$v_i^{(n)}<v_i^{(n)}{}'\le v_{i+1}^{(n)}$，粗略地有

$$Y_t\approx\sum_i v_i^{(n)}{}'[m(t,v_{i+1}^{(n)})-m(t,v_i^{(n)})]=\sum_{i,j} v_i^{(n)}m((s_i^{(n)},s_{i+1}^{(n)}]\times(v_j^{(n)},v_{j+1}^{(n)}]).$$

当$n\to\infty$时，可以证明上式右方按概率收敛．将此概率收敛的极限随机变量记为$\int_0^t\int_{-\infty}^{\infty} vm(ds,dv)$． 于是就得到下面的命题.

命题１７.２７ （用时空 Poisson 点过程表达非时齐复合 Poisson 过程Y_t的分解式）

$$Y_t = \int_0^t \int_{-\infty}^{\infty} vm(ds,dv).\qquad (17.21)'$$

3 过滤的 Poisson 过程

定义１７．２８ 设N_t是强度为 $l(t)$ 的非时齐的 Poisson 过程，$(X_n\}$ 是与之独立的独立同分布随机变量序列．又设 N_t 在 t 前第 n 个事件发生时刻 t_n 的赋值(或理解成能量)，不是像非时齐复合 Poisson 那样由 X_n 直接给出，而是通过某个有界的响应函数 $h(t,s,x)$ $(s \le t)$ 给出的 $h(t,t_n,X_n)$．于是累计能量为

$$Z_t = \int_0^t h(t,s,\tilde{X}_s)dN_s \qquad \left(\tilde{X}_s = \begin{cases} X_n & (s = t_n) \\ 0 & (\text{其它}) \end{cases}\right),$$

称之为过滤的 Poisson 过程．特别地，当$h(t,t_n,X_n)$ ＝X_n时，过滤的 Poisson 过程就是非时齐的复合 Poisson 过程．(所以过滤的 Poisson 过程相当于＂扭曲了的非时齐的复合 Poisson 过程＂)．

与命题１７．１９类似地可以证明

定义１７．２９ 过滤的 Poisson 过程Z_t在区间$[0,T]$上的特征泛函为： 对于任意连续的增函数 $F(t)$，定义

$$\overline{\Phi}_Z(F) \overset{\Delta}{=} Ee^{i\int_0^T Z_t dF(t)},\qquad (17.22)$$

(用 $\int Z_t dF(t)$ 形式的积分定义特征泛函要比用 $\int f(t)dZ_t$ 形式的积分更为灵活，使用面更广)．类似地还可以证明

命题１７.３０ 过滤的 Poisson 过程 Z_t 在区间$[0,T]$上的特征泛函有如下的表达式:

$$\overline{\Phi}_Z(F) = e^{\int_0^T [Ee^{i\int_u^T h(v,u,X)dF(v)} - 1])\, l(u)du}.\qquad (17.23)$$

并由此得到

$$EZ_t = \int_0^t E[h(t,u,X)]\, l(u)du,\quad Cov(Z_s, Z_t) = \int_0^{s\wedge t} E[h(s,u,X)h(t,u,X)]\, l(u)du.$$

用特征函数方法还不难证明下述类似于中心极限定理的近似定理：

定理１７.３１　(过滤的时齐的 Poisson 过程的泛函型的 Gauss 过程近似定理)

假定生成过滤的 Poisson 过程的 Poisson 过程是时齐的，其强度为 l． 那么，当 $l \to \infty$ 时，在 $(0,T]$ 上 $\dfrac{Z_t - EZ_t}{\sqrt{Var(Z_t)}}$ 按分布收敛到 x_t，其中 x_t 是期望函数为 0 的 Gauss 过程，其协方差函数为 $Cov(x_s, x_t) = r_{Z_s, Z_t}$ (指 Z_s 与 Z_t 的相关系数).

使用这个定理，可以将高强度的过滤的时齐的 Poisson 过程的近似计算大大地简化.

4　Poisson 随机微积分简介

４．１　关于时空 Poisson 点过程的随机积分

设 $\{m(t,v): t \geq 0, -\infty < v < \infty\}$ 是时空 Poisson 点过程，满足：对于任意 $s<t, a<b$ $m((s,t]\times(a,b]) \sim Poisson_{(l(t)-l(s))(F(b)-F(a))}$ (F 单调递增)．假定 $l(0)=0, F(\infty)-F(-\infty)=C<\infty$，那么，由 (S．1) 可知，计数过程 $N_t \overset{\Delta}{=} \int m(t,dv)$ 是强度为 $C\,l(t)$ 的非时齐的 Poisson 过程．我们不妨假定 $C=1$．

１．关于非时齐的 Poisson 过程 N_t 的随机积分

利用非时齐的 Poisson 过程在长度为 Δt 的区间上的增量大于 1 的概率为 $o(\Delta t)$，可以得到如下的引理.

引理１７．３２　对于非时齐的 Poisson 过程 N_t 及 $(0,t]$ 的一个划分：

$0 = t_0^{(n)} < \cdots < t_j^{(n)} < t_{j+1}^{(n)} < \cdots < t_n^{(n)} = t$，$\max_j (t_{j+1}^{(n)} - t_j^{(n)}) \to 0 (n\to\infty)$，我们有

$$\lim_{n\to\infty} \sum_{k=1}^{n} (N_{t_k^{(n)}} - N_{t_{k-1}^{(n)}})^2 = N_t. \qquad (17．24)$$

与 Brown 运动相比，这里是 N_t，而在 Brown 运动情形是 t．直观地就是

$$(dN_t)^2 = dN_t. \qquad (17．25)$$

证明　仿照 Brown 运动的相应的结论证明，细节留作习题.

定义１７．３３　对于 (N_t) 可知的，且轨道具有左极限的随机过程 Ψ_t，可以仿照 Ito 积分定义关于非时齐 Poisson 过程的随机积分如下:

$$\int_0^t \Psi_{s-} dN_s = \lim_{n\to\infty} \sum_{k=1}^{n} \Psi_{t_{k-1}^{(n)}} (N_{t_k^{(n)}} - N_{t_{k-1}^{(n)}}) . \qquad (17.26)$$

注意这里被积函数用的是左极限，原因在于积分和中被积函数取的是区间左端点的值.

例１７.３４

$$\int_0^t N_{s-} dN_s = \frac{1}{2} N_t^2 - \frac{1}{2} N_t .$$

证明　仿照 Brown 运动情形，由定义直接算得.

关于非时齐 Poisson 过程的随机积分有类似于 Ito 积分的许多性质，如可加性，以及

（1）$\int_0^t \Psi_{s-} dN_s - \int_0^t \Psi_s l(s) ds$ 是鞅.

(2) $(\int_0^t \Psi_{s-} dN_s)^2 - \int_0^t \Psi_s^2 l(s) ds$ 也是鞅.

更一般地，我们可以定义

２．关于时空 Poisson 点过程 $m(t,v)$ 的随机积分

对于带有参数 v 的 $(m(t,v):\forall v)$ 可知的，且对 t 有左极限的随机过程 $\Psi_t(v)$ 及如下的划分：$\cdots < v_i^{(n)} < v_{i+1}^{(n)} < \cdots < v_{m_n}^{(n)}, v_0^{(n)} = -n, v_{m_n}^{(n)} = n$，$0 = s_0^{(n)} < \cdots < s_j^{(n)} < s_{j+1}^{(n)} < \cdots < s_n^{(n)} = t$，$\max_{i,j}(v_{i+1}^{(n)} - v_i^{(n)}, s_{j+1}^{(n)} - s_j^{(n)}) \to 0 (n \to \infty)$，$v_i^{(n)} < v_i^{(n)}{}' \le v_{i+1}^{(n)}$，定义关于时空 Poisson 点过程的随机积分为

$$\int_0^t \int \Psi_{s-}(v) m(ds, dv) = \lim_{n\to\infty} \sum_{i,j} \Psi_{s_{j-1}^{(n)}}(v_i^{(n)}) m((s_i^{(n)}, s_{i+1}^{(n)}] \times (v_j^{(n)}, v_{j+1}^{(n)}]) . (17.27)$$

类似地，还有

$$(m(dt,dv))^2 = m(dt,dv) \qquad (17.25)'$$

这种随机积分有类似于 Ito 积分的许多性质，如可加性，以及

（1）$\int_0^t \int \Psi_{s-}(v) m(ds,dv) - \int_0^t \int \Psi_s(v) l(s) ds dF(v)$ 是鞅.

（2）$[\int_0^t \int \Psi_{s-}(v) m(ds,dv)]^2 - \int_0^t \int \Psi_s^2(v) l(s) ds dF(v)$ 也是鞅.

设对应于 $N_t \overset{\Delta}{=} \int m(t,dv)$ 的事件发生时刻列为 $\{t_n\}$．由于 F 是分布函数，此时空 Poisson 点过程由非时齐的复合 Poisson 过程生成.设此非时齐的复合 Poisson 过程在 t_n 的

赋值为 X_n，于是我们有

$$\int_0^t \int \Psi_{s-}(v)m(ds,dv) = I_{\{N_t \ge 1\}} \sum_{n=1}^{N_t} \Psi_{t_n}(X_n) .$$

４．２　以 Poisson 过程或以时空 Poisson 点过程驱动的随机微分方程与 Poisson 随机微积分的复合函数的 Ito 公式

时空 Poisson 过程驱动的随机积分方程

$$x_t = x_0 + \int_0^t b(s,x_s)ds + \int_0^t \int g(s,x_{s-},v)m(ds,dv) \qquad （１７．２８）$$

也称为时空 Poisson 过程驱动的随机微分方程，其中$b(t,x), g(t,x,v)$是(t,x)的连续函数.

（１７．２８）也常常简记为如下的微分形式

$$dx_t = b(t,x_t)dt + \int g(t,x_{t-},v)m(dt,dv) . \qquad （１７．２８）'$$

而由非时齐的 Poisson 过程驱动的随机微分方程

$$x_t = x_0 + \int_0^t b(s,x_s)ds + \int_0^t g(s,x_{s-},v)dN_s \qquad （１７．２９）$$

就正是如下的特殊形式

$$x_t = x_0 + \int_0^t b(s,x_s)ds + \int_0^t \int g(s,x_{s-},v)m(ds,dv) ,$$

即（１７．２９）是（１７．２８）的特例. 如果一个随机过程x_t满足（１７．２８），就称为方程（１７．２８）的解.　这时有

$$x_t - x_{t-} = \int g(t,x_{t-},v)m(\{t\},dv) ,$$

未必一定等于 0. 所以（１７．２８）的解的轨道与 Ito 方程的解的轨道很不同，Ito 方程的解是连续的，而是（１７．２８）的解是可以间断的. 此外，与 Ito 方程完全类似地，可以证明时空 Poisson 过程驱动的随机微分方程的解也是 Markov 过程. 于是时空 Poisson 过程驱动的随机微分方程就成为一类轨道可以有间断的 Markov 过程的建模的有力工具.例如，可以用它改进风险证券的 Black-Scholes 模型，使之可以包容市场的突然变化.

定义１７。３５　(Poisson 型 Ito－Skorohod 过程)

形如

$$dx_t = \Phi_t dt + \int \Psi_{t-}(v)m(dt,dv) \qquad （１７．３０）$$

的随机过程x_t称为 Poisson 型的 Ito-Skorohod 过程(只对v作积分).

与 Ito 过程类似，Poisson 型的 Ito-Skorohod 过程对于连续可微的复合运算也是封闭的. 这就是下面的链法则.

定理１７.３６　(时空 Poisson 过程驱动的随机微分方程的解的复合函数－Ito 公式)

若$F(t,x)$连续可微，x_t是用（１７.３０）表示的 Poisson 型的 Ito-Skorohod 过程，

则 $F(t,x_t)$ 也是 Poisson 型 Ito-Skorohod 过程，且其表示为

$$dF(t,x_t)=(F_t'+\Phi_t F_x')dt+\int[F(t,x_{t-}+\Psi_{t-}(v))-F(t,x_{t-})]m(dt,dv).$$

特别，如果 x_t 是随机微分方程（１７．２８）的解，则有

$$dF(t,x_t)=(F_t'+b(t,x_t)F_x')dt+\int[F(t,x_{t-}+g(t,x_{t-},v))-F(t,x_{t-})]m(dt,dv). \tag{１７．３１}$$

直观证明　要点是在将 $(0,t]$ 细分后，在区间集 $\{(t_{j-1}^{(n)},t_j^{(n)}]\}$ 中有的包含一个 t_k 的区间,我们将它改记为 $(t_{j-1,k}^{(n)},t_{j,k}^{(n)}]$．而有的区间不包含 $\{t_k\}$ 中的任一个，对于这类区间我们有 $\int_{t_{j-1}^{(n)}}^{t_j^{(n)}}\Psi_{s-}(v)m(ds,dv)=0$．所以，当 n 充分大时有

$$\sum_{j=1}^{n}[F(t_{j-i}^{(n)},x_{t_{j-1}^{(n)}}+\int_{t_{j-1}^{(n)}}^{t_j^{(n)}}\Psi_{s-}(v)m(ds,dv))-F(t_{j-1}^{(n)},x_{t_{j-1}^{(n)}})]$$

$$\approx I_{\{N_t\ge1\}}\sum_{k=1}^{N_t}[F(t_{j-i,k}^{(n)},x_{t_{j-1,k}^{(n)}}+\int_{t_{j-1,k}^{(n)}}^{t_{j,k}^{(n)}}\Psi_{s-}(v)m(ds,dv))-F(t_{j-1,k}^{(n)},x_{t_{j-1,k}^{(n)}})]$$

$$\xrightarrow{n\to\infty}I_{\{N_t\ge1\}}\sum_{k=1}^{N_t}[F(t_k,x_{t_k}+\Psi_{t_k}(v)m(ds,dv))-F(t_k,x_{t_k})]$$

$$=\int_0^t\int[F(t,x_{t-}+\Psi_{t-}(v))-F(t,x_{t-})]m(dt,dv).$$

例１７．３７　若 x_t 是如下的 Poisson 型的 Ito-Skorohod 过程:

$$dx_t=\Phi_t dt+\int\Psi_{t-}(v)m(dt,dv),$$

则有

（1）　幂：

$$dx_t^m=mx_t^{m-1}\Phi_t dt+\int[(x_{t-}+\Psi_{t-}(v))^m-x_t^m]m(dt,dv). \tag{１７．３２}$$

（2）　指数：　令　$h_t=e^{x_t}$，则

$$dh_t=h_t\Phi_t dt+\int h_t(e^{\Psi_{t-}(v)}-1)m(dt,dv),$$

即 h_t 满足如下的随机微分方程:

$$dh_t=h_t dx_t+\int h_t(e^{\Psi_{t-}(v)}-\Psi_{t-}(v)-1)m(dt,dv). \tag{17．32}$$

（3） 对数： 如果还有$x_t > 0$，记 $V_t = \ln x_t$，则

$$dV_t = \frac{\Phi_t}{x_t}dt + \int \ln(1+\frac{\Psi_{t-}(v)}{x_t})m(dt,dv) .\qquad（１７．３４）$$

对于d维的情形，即x_t取值于R^d的情形类似的结论仍然正确．

再则，时空 Poisson 点过程也可以是多维的，例如，是由多维的非时齐的复合 Poisson 过程

$\vec{Y}_t = \sum_{n=1}^{N_t} \vec{X}_n$所导出的．

例１７．３８ （乘积的 Ito 公式与分部积分） 对于随机微分方程

$$d\vec{x}_t = \int \vec{g}(t,\vec{x}_{t-},\vec{v})m(dt,d\vec{v})$$

及$F(t,\vec{x})$有

$$dF(t,\vec{x}_t) = \int [F(t,\vec{x}_{t-} + \vec{g}(t,\vec{x}_{t-},\vec{v})) - F(t,\vec{x}_{t-})]m(dt,d\vec{v}) .$$

特别地，对于

$$dx_t^{(i)} = \int g_i(t,x_{t-}^{(i)},\vec{v})m(dt,d\vec{v}) \qquad (i=1,2)$$

有乘积公式

$$d(x_t^{(1)}x_t^{(2)}) = \int [x_t^{(1)}g_2(t,x_{t-}^{(2)},v) + x_t^{(2)}g_1(t,x_{t-}^{(1)},v) +$$

$$+ g_1(t,x_{t-}^{(1)},v)g_2(t,x_{t-}^{(2)},v)]m(dt,dv) .$$

即

$$d(x_t^{(1)}x_t^{(2)}) = x_t^{(1)}dx_t^{(2)} + x_t^{(2)}dx_t^{(1)} + \int g_1(t,x_{t-}^{(1)},v)g_2(t,x_{t-}^{(2)},v)m(dt,dv) . \qquad (17.35)$$

（１７．３５）可以推广到一般情形，即

$$d\vec{x}_t = \vec{b}(t,\vec{x}_t)dt + \int \vec{g}(t,\vec{x}_{t-},\vec{v})m(dt,d\vec{v})$$

的情形仍然正确．从而得到一般的分部积分公式．请读者自己写下这个公式．

下面的存在唯一性的证明,可以在随机微分方程的书的有关带跳的随机微分方程的章节中找到，我们略去其证明．

定理１７。３９ （时空 Poisson 过程驱动的随机微分方程解的存在唯一性）

若$l(t)$有界，$g(t,x,v)$连续，且在$[0,T]$上满足

$$b(t,x)^2 + g(t,x,v)^2 \le C(1+x^2) ,\qquad（１７．３６）$$

$$|b(t,x) - b(t,y)| + |g(t,x,v) - g(t,y,v)|^2 \le C|x-y|^2 .\qquad（１７．３７）$$

那么随机微分方程（１７.２８）存在唯一的解x_t，而且它是 Markov 过程．

例１７.４０ 设$f(t)$是数值函数，N_t是非时齐的 Poisson 过程．考虑随机微分方程

$$dV_t = V_t f(t) dN_t, V_0 = 1.$$

它有唯一解．可以用 Ito 公式验证此解正是（利用例１７.３７（２））

$$V_t = e^{\int_0^t \ln(1+f(s))dN_s} = \prod_{n=1}^{N_t} (1+f(t_n)). \qquad （１７．３８）$$

４．３ 由 Brown 运动和时空 Poisson 过程联合驱动的随机微分方程

设 Brown 运动$\{B_t : t \geq 0\}$与时空 Poisson 点过程$\{m(t,v): t \geq 0, 一切 v\}$相互独立．则由 Brown 运动和时空 Poisson 过程联合驱动的随机微分方程

$$x_t = x_0 + \int_0^t b(s,x_s)ds + \int_0^t s(s,x_s)dB_s + \int_0^t \int g(s,x_{s-},v)m(ds,dv), \qquad (17.39)$$

即方程

$$dx_t = b(t,x_t)dt + s(t,x_t)dB_t + \int g(t,x_{t-},v)m(dt,dv)$$

的解是 Markov 过程，我们给出如下的两个定理，而略去其证明．

定理１７。４１ （解的存在唯一性） 若$l(t)$有界，$g(t,x,v)$连续，且在$[0,T]$上满足

$$b(t,x)^2 + g(t,x,v)^2 + s(t,x)^2 \leq C(1+x^2), \qquad (17.40)$$

$$|b(t,x)-b(t,y)| + |s(t,x)-s(t,y)|^2 + |g(t,x,v)-g(t,y,v)|^2 \leq C|x-y|^2. \qquad (17.41)$$

那么随机微分方程（１７．３９）存在唯一的解x_t，而且它是 Markov 过程．

定理１７．４２ （一般的 Ito 公式）

若$F(t,x)$对t连续可微，对x二阶连续可微，x_t是（１７.３９）的解，则

$$dF(t,x_{t-}) = [F_t' + b(t,x_t)F_x' + \frac{1}{2}s^2(t,x_t)F_{xx}'']dt + s(t,x_t)F_x' dB_t$$
$$+ \int [F(t,x_{t-}+g(t,x_{t-},v)) - F(t,x_{t-})]m(dt,dv).$$

同样，在多维情形也有对应的定理．

例１７.４３ 随机微分方程

$$dV_t = V_t(g(t)dB_t + f(t)dN_t), V_0 = 1$$

的唯一解是

$$V_t = e^{\int_0^t g(s)B_s - \frac{1}{2}\int_0^t g(s)^2 ds} \prod_{n=1}^{N_t} (1+f(t_n)). \qquad （１７．４２）$$

5　自激点过程

5．1　自激点过程的强度过程与条件计数强度

定义１７．４４　如果计数过程N_t满足：$N_0=0$，且当$h\to 0$时

$$P(N_{t+h}-N_t=1\,|\,N_t=0)=l(t,0)h+o(h)，$$

$$P(N_{t+h}-N_t=1\,|\,N_t=k,t_1=s_1,\cdots,t_k=s_k)=l(t,N_t;s_1,\cdots,s_k)h+o(h)$$

$(k\ge 1)$，

$$P(N_{t+h}-N_t\ge 2\,|\,N_t=k,t_1=s_1,\cdots,t_k=s_k)=o(h)\quad (k\ge 1).\qquad（１７．４３）$$

则称为自激点过程．而随机过程

$$l_t\overset{\Delta}{=}\begin{cases} l(t,0) & (t<t_1) \\ l(t,N_t;t_1,\cdots,t_{N_t}) & (t_{N_t}\le t<t_{N_t+1}\} \end{cases},\qquad（１７．４４）$$

称为其强度(随机)过程．

最简单的是 $l_t=l(t,N_t)$ 的情形．

定义１７.４５

$$\hat{l}(t,N_t)=E(l_t\,|\,N_t)\qquad（１７．４５）$$

称为自激点过程的条件计数强度．

于是我们有

$$P(N_{t+h}-N_t=1\,|\,N_t)=\hat{l}(t,N_t)h+o(h)，\ P(N_{t+h}-N_t\ge 2\,|\,N_t)=+o(h)$$

（１７．４６）

例１７．４６　在第１３章第５节中，年龄x的n个人的群体在时刻t以前的死亡人数（死亡计数过程）N_t是自激点过程，其强度过程为 $l(t,N_t)=(n-N_t)m_{x+t}$，其中 m_{x+t} 是年龄为$x+t$的死亡率．

5．2　自激点过程的绝对概率

仿照非时齐的 Poisson 过程的绝对概率的推导，我们可以得到

$$P(N_t=0\,|\,N_s=0)=e^{-\int_s^t l(u,0)du}，$$

$$P(N_t-N_s=0\,|\,N_s,t_1,\cdots,t_{N_s})=e^{-\int_s^t l(u,N_s;t_1,\cdots,t_{N_s})du}，\ (N_s\ge 1).$$

(17. 47)

令

$$p_n(t)=P(N_t=n).\qquad (17.\ 48)$$

同样地仿照非时齐的 Poisson 过程的绝对概率的推导，利用（１７.４６）我们得到

定理１７．４７　自激点过程的绝对概率满足如下的无穷个常微分方程

$$p_0'(t) = -\hat{l}(t,0)\,p_0(t)\text{，}$$

$$p_n'(t) = -\hat{l}(t,n)\,p_n(t) + \hat{l}(t,n-1)\,p_{n-1}(t) \quad (n \ge 1)\text{．}$$

这是一个非时齐的单侧纯生过程．可以用归纳法求得其解为

$$p_0(t) = e^{-\int_0^t \hat{l}(u,0)du} = e^{-\int_s^t l(u,0)du}\text{，}$$

$$p_n(t) = \int_0^t \hat{l}(s,n-1)\,p_{n-1}(s)e^{-\int_s^t \hat{l}(u,n)du}ds \quad (n \ge 1)\text{．} \qquad (17.49)$$

５．３　自激点过程的事件到达时刻的联合分布

我们求自激点过程 N_t 的前 n 个事件到达时刻 $(t_1,\cdots,t_n)$ 的联合分布．注意

$$l(t,n;s_1,\cdots,s_n)h + o(h) = P(N_{t+h} - N_t = 1 \mid N_t = n, t_1 = s_1,\cdots,t_n = s_n)$$

$$= \frac{P(t < t_{n+1} \le t+h, N_t = n \mid t_1 = s_1,\cdots,t_n = s_n)}{P(N_t = n \mid t_1 = s_1,\cdots,t_n = s_n)}$$

$$= \frac{P(t < t_{n+1} \le t+h \mid t_1 = s_1,\cdots,t_n = s_n)}{P(t_{n+1} > t \mid t_1 = s_1,\cdots,t_n = s_n)}\text{．}$$

这正说明了下述结论：

定理１７.４８　$l(t,n;s_1,\cdots,s_n)$ 是在 $t_1 = s_1,\cdots,t_n = s_n$ 的条件下，t_{n+1} 的故障率．因此，在 $t_1 = s_1,\cdots,t_n = s_n$ 的条件下，t_{n+1} 的条件密度为

$$p_{t_{n+1}|t_1,\cdots,t_n}(s_{n+1} \mid s_1,\cdots,s_n) = e^{-\int_{s_n}^{s_{n+1}} l(u,n;s_1,\cdots,s_n)du}\text{．（１７．５０）}$$

从而 $(t_1,\cdots,t_n)$ 的联合分布为

$$p_{t_1,\cdots,t_n}(s_1,\cdots,s_n) = \prod_{k=1}^{n} e^{-\int_{s_{k-1}}^{s_k} l(u,k-1;s_1,\cdots,s_{k-1})du}\text{，}\qquad\text{（１７．５１）}$$

其中 $s_0 = 0,\ l(t,0;s_0) = l(t,0)$．

我们可以进一步由（１７．５１），求得自激点过程在时刻 t 时的计数 N_t 与事件发生时刻的联合分布 $P(N_t = n, t_1 \le s_1,\cdots,t_n \le s_n)$ 关于 $(s_1,\cdots,s_n)$ 的密度（称之为自激点过程

的样本分布）

$$p_{N_t,t_1,\cdots,t_{N_t}}(n,s_1,\cdots,s_n)=P(N_t=n\,|\,t_1=s_1,\cdots,t_n=s_n)\ p_{t_1,\cdots,t_n}(s_1,\cdots,s_n)$$

$$=P(N_t-N_{s_n}=0\,|\,t_1=s_1,\cdots,t_n=s_n)\prod_{k=1}^{n}e^{-\int_{s_{k-1}}^{s_k}l(u,k-1;s_1,\cdots,s_{k-1})du}$$

$$=e^{-\int_{s_n}^{t}l(u,n;s_1,\cdots,s_n)du}\prod_{k=1}^{n}e^{-\int_{s_{k-1}}^{s_k}l(u,k-1;s_1,\cdots,s_{k-1})du}.$$

此即下面的定理.

定理１７．４９　自激点过程计数N_t与事件发生时刻的联合分布（即自激点过程的样本分布）$P(N_t=n,t_1\le s_1,\cdots,t_n\le s_n)$关于$(s_1,\cdots,s_n)$的密度为（$n\ge 1$）

$$p_{N_t,t_1,\cdots,t_{N_t}}(n,s_1,\cdots,s_n)=e^{-\int_{s_n}^{t}l(u,n;s_1,\cdots,s_n)du}\prod_{k=1}^{n}e^{-\int_{s_{k-1}}^{s_k}l(u,k-1;s_1,\cdots,s_{k-1})du}\quad.（１７．５２）$$

例１７．５０　　伽玛光子检测器用以检测按强度函数为$n(t)$的非时齐的 Poisson 过程到达的伽玛光子．但是，这种装置在检测到一个光子后有一个随机的失效时间．假定检测到第n个光子后的失效时间为V_n，且$\{V_n\}$独立同分布，并与此非时齐的 Poisson 过程独立，其共同分布函数是$F_V(t)$．将至时刻t为止检测到的光子数记为N_t，再记各个检测到光子的时刻分别依次为　$t_1,\cdots,t_n,\cdots$．

假定开始时检测装置是有效的．我们说明N_t是自激点过程．事实上，此时有$l(t,0)=n(t)$．又因为在时刻t_n检测到光子后，在其后的时刻t检测装置不失效的概率是$F_V(t-t_n-0)$（此处"-0"表示左极限)．于是对于$N_t=n\ge 1$有

$$l(t,n;s_1,\cdots,s_n)=n(t)F_V(t-t_n-0)\ （l(t,N_t;t_1,\cdots,t_{N_t})=n(t)F_V(t-t_{N_t}-0)）.$$

特别当$V_n=$常数的时候，各个检测到光子的时刻　$(t_1,\cdots,t_n)$的联合分布可以容易地由（１７．５１）计算得到.

５．４　具有限记忆的自激点过程

定义１７．５１　（０－记忆自激点过程）自激点过程N_t称为０－记忆的，如果

$$l(t,N_t;t_1,\cdots,t_{N_t})=l(t,N_t)$$

（即与$t_1,\cdots,t_{N_t}$无关）.

例１７．４６中的自激点过程就是０－记忆的.

命题１７．５２　自激点过程N_t是 Markov 链，当且仅当它是０－记忆的.

证明　必要性只需验证，对于$s > s_m > \cdots > s_1$而言，$P(N_{s+t} = j \mid N_s = i)$与$P(N_{s+t} = j \mid N_s = i, N_{s_m} = i_m, \cdots, N_{s_1} = i_1)$满足同样的方程.　】

定义１７．５３　（m－记忆自激点过程）　自激点过程N_t称为m－记忆的，如果

$$l(t, N_t; t_1, \cdots, t_{N_t}) = l(t, N_t; t_{N_t - m+1}, \cdots, t_{N_t})$$

（即与$t_1,\cdots,t_{N_t - m}$无关）.

命题１７．５４　自激点过程N_t是m－记忆的，当且仅当$\{t_n\}$是m阶 Markov 链，即 $\dot{x}_n \overset{\Delta}{=} \begin{pmatrix} t_{n+1} \\ \\ t_{n+m} \end{pmatrix}$是 Markov 链.

证明　由（１７．５０）即得.

定义１７.５５　（时齐的１－记忆自激点过程）　１－记忆自激点过程N_t称为时齐的，如果存在一个函数$h(x, y)$使

$$l(t, N_t; t_{N_t}) = h(N_t, t - t_{N_t}).$$

由（１７．５０）立刻得到下面的定理

定理１７.５６　设１－记忆自激点过程N_t，其事件发生的时刻分别依次为$\{t_n\}$. 那么

（１）　N_t是时齐的充要条件是$\{t_n\}$是独立随机变量序列的部分和.

（２）$\{t_n\}$是延迟更新流的充要条件是，存在一个函数$g(y)$使

$$l(t, N_t; t_{N_t}) = g(t - t_{N_t}).$$

（３）$\{t_n\}$是更新流的充要条件是，存在一个函数$h(t)$使

$$l(t, 0) = h(t)，且\ l(t, N_t; t_{N_t}) = h(t - t_{N_t}),$$

其中$h(t)$可以解释为待更新的部件的故障率.

５．５　对于自激点过程的随机积分

设 N_t 是自激点过程，又随机过程 Ψ_t 是 (N_t) 可知的，有界的，轨道为分段连续的随机过程．那么可以如关于非时齐的 Poisson 的随机积分类似地定义

$$\int_0^t \Psi_{u-} dN_u = \begin{cases} 0 & (N_t = 0) \\ \sum_{n=1}^{N_t} \Psi_{t_n} & (N_t \geq 1) \end{cases}.$$

此时有

$$E[\int_0^t \Psi_{u-} dN_u] = \int_0^t E[\Psi_u l_u] du.$$

对于这种随机积分，同样可以讨论随机微分方程及 Ito 公式，其形式是一样的．

例１７．５７　设

$$dx_t = -l_t dt + \ln l_t dN_t,$$

那么由（１７．９）可知其指数函数 $h_t = e^{x_t}$ 可以看成自激点过程 N_t 的似然函数，由 Ito 公式得到，它的 n 次幂满足

$$dh_t^n = -nh_t^n l_t dt + h_t^n (l_t^n - 1) dN_t.$$

[注]　还可以抽象出更为一般的时空点过程及其随机微积分．

５．６　二重 Poisson 过程

定义１７．５８　设 x_t 是一个随机过程，一般地称为被调制的信息过程．而当 x_t 的轨道给定时，计数过程 N_t 是一个强度函数为 $l(t,x_t)$ 的非时齐的 Poisson 过程，那么 N_t 称为二重 Poisson 过程．$l(t,x_t)$ 称为二重 Poisson 过程的强度．

在实用问题中，常常是通过接收到的点过程 N_t 的一段样本轨道，来估计被调制的信息过程．例如，信号为函数 $S(t)$，它在被均值为０的背景 Gauss 过程噪声的干扰下，成为代表信息的 Gauss 过程 x_t，而由此生成的二重 Poisson 过程 N_t 则是光子流．最简单的情形是强度为 $l(t,x_t) = xg(t)$（随机变量乘以常函数），$l(t,x_t) = cx_t^2$．

这时有

$$p_n(t) \stackrel{\Delta}{=} P(N_t = n) = E[e^{-\int_0^t l(s,x_s)ds} \frac{[\int_0^t l(s,x_s)ds]^n}{n!}].$$

定理１７．５９　假定二重 Poisson 过程 N_t 的强度 $l(t,x_t)$ 的数学期望有限．记 N_t 的

事件列分别依次为$0<t_1<\cdots<t_n<\cdots$．令

$$l_t=E(l(t,x_t)\mid N_u:u\le t)=\begin{cases}E(l(t,x_t)\mid N_t) & (N_t=0)\\ E(l(t,x_t)\mid N_t,t_1,\cdots,t_{N_t}) & (N_t\ge 1)\end{cases}.$$

那么，N_t是强度随机过程为l_t的自激点过程．

证明　当$h\to 0$时

$$E[P(N_{t+h}-N_t=1\mid N_s:s\le t)=E[P(N_{t+h}-N_t=1\mid N_s,:s\le t,x_t)\mid N_s:s\le t)$$

$$=E[E(l(t,x_t)h\mid N_s:s\le t)\mid N_s:s\le t]+o(h)=l_t h+o(h).$$

同样证明

$$P(N_{t+h}-N_t\ge 2\mid N_s:s\le t)=o(h).$$　】

于是二重 Poisson 过程的一些统计量的计算就直接化归自激点过程的计算．定理１７．４７，定理１７．４８，定理１７．４９也都成立．

［注］一般由于$\hat{l}(t,n)$不易计算，所以$p_n(t)$也不易计算．下面的定理可用于近似算法，它可以用特征函数方法证明

定理１７．６０　（二重 Poisson 过程绝对概率的近似计算）　对二重 Poisson 过程N_t，记

$$\Lambda_t=\int_0^t l(s,x_s)ds,\quad N_t^*=\frac{N_t-EN_t}{\sqrt{Var(N_t)}},\Lambda_t^*=\frac{\Lambda_t-E\Lambda_t}{\sqrt{Var(\Lambda_t)}}.$$

假定$t\to\infty$时，$E\Lambda_t\to\infty$．那么

（１）　若还有$\dfrac{Var(\Lambda_t)}{(E\Lambda_t)^2}\to c>0$，则只要$\Lambda_t^*$按分布收敛于随机变量$h$，$N_t^*$就按分布收敛到$\dfrac{x}{\sqrt{1+c}}+\dfrac{h}{\sqrt{1+\frac{1}{c}}}$，其中$x$与$h$独立，且$x\sim N(0,1)$．

（２）　若还有$\dfrac{Var(\Lambda_t)}{(E\Lambda_t)^2}\to 0$，则$N_t^*$按分布收敛$x$，其中$x\sim N(0,1)$．

（３）　若还有$\dfrac{Var(\Lambda_t)}{(E\Lambda_t)^2}\to\infty$，则只要$\Lambda_t^*$按分布收敛于随机变量$h$，$N_t^*$也按分布收敛于$h$．

定义１７．６１　设x_t是一个随机过程（可以是多维的），一般地称为被调制的信息过程．而当x_t的轨道给定时，计数过程$\vec{N}_t=(N_t^{(1)},\cdots,N_t^{(d)})$的各个分量彼此独立，且$N_t^{(k)}$

是一个强度函数为 $l^{(k)}(t,x_t)$ 的非时齐的 Poisson 过程，那么 $\overrightarrow{N}_t$ 称为多道二重 Poisson 过程．$\overrightarrow{l}(t,x_t)=(l_t^{(1)}(t,x_t),\cdots,l_t^{(d)}(t,x_t))$ 称为多道二重 Poisson 过程的强度．

在应用中，如核医疗中多探针的检测系统，伽玛射线照相机等，常用 d 道二重 Poisson 过程建模．更多地出现的是 x_t 是 Markov 过程，或 Markov 链的情形．用 $\overrightarrow{N}_t$ 的观测样本去估计信息过程 x_t 就需要滤波．

习题　17

1．证明非时齐的复合 Poisson 过程是非时齐的独立增量过程．再求它的数学期望函数与协方差函数．

2． 如果赋值随机变量具有分布 $X_n \sim \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$．问这时的非时齐的复合 Poisson 过程 是什么？解释其概率含义．

3． 求 $\lim_{n\to\infty}\sum_{k=1}^{n} N_{t_k^{(n)}}(N_{t_k^{(n)}}-N_{t_{k-1}^{(n)}})$．

4．若 $dx_t=\Phi_t dt+\int \Psi_{t-}(v)m(dt,dv)$，对于复指数 $h_t=e^{ix_t}$，求 dh_t．

5．设 $f(t)$ 是数值函数，N_t 是非时齐的 Poisson 过程．V_t 满足随机微分方程

$$dV_t=V_t f(t)dN_t, V_0=1.$$

（1） 证明 $EV_t=e^{\int_0^t f(s)l(s)ds}$．

（2） 证明 $dV_t^2=V_t^2[2+f(t)]f(t)dN_t$．

（3） 求 $Var(V_t)$，$Var(V_t^2)$．

6．即用定理１７．４７，证明自激点过程 N_t 的数学期望和方差分别为

$$EN_t=\int_0^t (El_u)du,$$

$$Var(N_t)=\int_0^t (El_u)du+2\int_0^t E(N_u l_u)du-[\int_0^t (El_u)du]^2,$$

其中随机过程 l_t 是 N_t 的随机强度过程（见（１７．４４)．

7．设 N_t 是非时齐的生灭过程，即０－记忆自激点过程，其 $l(t,N_t)$ 满足 $\sum_n \frac{1}{\sup_{s\le t} l(s,n)} = \infty$，

证明 $P(N_t < \infty) = 1$．　（提示　应用定理１７．４７）．

8．设０－记忆自激点过程 N_t 满足： $l(t,N_t) = l(t,0)[N_t + 1]$．证明 $p_n(t) = p_0(t)(1 - p_0(t))^n$．

9．设二重 Poisson 过程 N_t 的 $l(t,x_t) = g(t)x$，其中 $g(t)$ 非随机，　x 有分布密度 $p_x(t)I_{[0,\infty)}(t)$．

证明

$$P(N_t = n) = \frac{[\int_0^t g(s)ds]^{nt}}{n!}\int_0^\infty t^n e^{-t\int_0^t g(s)ds} p_x(t)dt\,.$$

再证明

$$\hat{l}(t,N_t) = E(l_t \mid N_t) = E[E(l(t,x_t) \mid N_s : s \le t) \mid N_t]$$

$$= E[l(t,x_t) \mid N_t] = \frac{g(t)\int_0^\infty s^{N_t+1} e^{-s\int_0^t g(u)du} p_x(s)ds}{\int_0^\infty s^{N_t} e^{-s\int_0^t g(u)du} p_x(s)ds}\,.$$

最后，如果 x ~ Gamma 分布 $\Gamma(a, l)$，那么

$$P(N_t = n) = \frac{\Gamma(n+a)}{n!\Gamma(a)}\left[\frac{l}{l + \int_0^t g(s)ds}\right]^a \left[\frac{\int_0^t g(s)ds}{l + \int_0^t g(s)ds}\right]^n,$$

N_t 称为非时齐的 Polya 过程．　证明若 $x \sim \exp_l$，则（１７．９）的似然函数为

$$p_{N_t,t_1,\cdots,t_{N_t}}(N_t, t_1, \cdots, t_n) = (N_t)!\, l[l + \int_0^t g(s)ds]^{-(N_t+1)} e^{\int_0^t \ln l(s)dN_s}\,.$$

并证明建模参数 l 的最大似然估计是

$$\hat{l} = \frac{1}{N_t}\int_0^t g(s)ds\,.$$

10．对于二重 Poisson 过程 N_t，记 $\Lambda_t = \int_0^t l(s,x_s)ds$．证明

$$Cov(N_s, N_t) = Cov(\Lambda_s, \Lambda_t) + E\Lambda_{s\wedge t}\,.$$

11．对于非时齐的 Poisson 过程 N_t 及 $(0,t]$ 的一个划分：

$$0=t_0^{(n)}<\cdots<t_j^{(n)}<t_{j+1}^{(n)}<\cdots<t_n^{(n)}=t\text{，}\max_j(t_{j+1}^{(n)}-t_j^{(n)})\to 0(n\to\infty)\text{，}$$

证明 $\lim_{n\to\infty}\sum_{k=1}^{n}(N_{t_k^{(n)}}-N_{t_{k-1}^{(n)}})^2=N_t$．